

# FLEXIBLE MODELS OF TIME-VARYING EXPOSURES

Chenkun Wang

Submitted to the faculty of the University Graduate School  
in partial fulfillment of the requirements  
for the degree  
Doctor of Philosophy  
in the Department of Biostatistics,  
Indiana University

May 2015

Accepted by the Graduate Faculty, Indiana University, in partial  
fulfillment of the requirements for the degree of Doctor of Philosophy.

---

Sujuan Gao, Ph.D., Co-Chair

Doctoral Committee

---

Hai Liu, Ph.D., Co-Chair

---

Zhangsheng Yu, Ph.D.

March 27, 2015

---

Christopher M. Callahan, M.D.

© 2015

Chenkun Wang

## DEDICATION

*To My Parents and My Husband*

## ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor, Dr. Sujuan Gao, and co-advisor, Dr. Hai Liu, for their great teaching, enduring encouragement, and endless patience. Their true enthusiasm, broad knowledge and unique insights in statistics inspired me tremendously during my graduate school years. They demonstrated to me the importance of biostatistics and showed me the way to make significant contributions to the society by conducting biostatistics research that is critical in our fighting with diseases. Their guidance will accompany me along my journey in biostatistics and I will continue to benefit from it. I deeply appreciate their supports and contributions to my research, without which this work would have been impossible.

I also want to extend my thanks to Dr. Zhangsheng Yu and Dr. Christopher M. Callahan, for their extensive knowledge, insightful comments to my thesis and for their assistance at every aspect of my research. In addition, I would like to acknowledge the fellowship from School of Science for my first year and two NIH grants, R01 AG019181 and P30 AG10133, for my research assistantship.

It has been wonderful to study and work in Department of Biostatistics Indiana University. I would like to give my special thanks to Chair, Dr. Barry Katz and all the other faculties, who commit to training junior people professionally. In particular, the collaboration with the Regenstrief Institute has been a truly precious opportunity for me, where I developed collaboration skills, I learned how to apply my knowledge to applications and, most importantly, I obtained experience of solving real-world problems. I also feel thankful for daily help from all the current and former students on biostatistics and beyond biostatistics.

Finally, my very special thanks go to my beloved husband, Wenqin Xu, for his support and companion. He makes me feel strong and bravely face every challenge. I must also thank my parents for their unconditional love and support through my life.

Chenkun Wang

## FLEXIBLE MODELS OF TIME-VARYING EXPOSURES

With the availability of electronic medical records, medication dispensing data offers an unprecedented opportunity for researchers to explore complex relationships among long-term medication use, disease progression and potential side-effects in large patient populations. However, these data also pose challenges to existing statistical models because both medication exposure status and its intensity vary over time. This dissertation focused on flexible models to investigate the association between time-varying exposures and different types of outcomes. First, a penalized functional regression model was developed to estimate the effect of time-varying exposures on multivariate longitudinal outcomes. Second, for survival outcomes, a regression spline based model was proposed in the Cox proportional hazards (PH) framework to compare disease risk among different types of time-varying exposures. Finally, a penalized spline based Cox PH model with functional interaction terms was developed to estimate interaction effect between multiple medication classes. Data from a primary care patient cohort are used to illustrate the proposed approaches in determining the association between antidepressant use and various outcomes.

Sujuan Gao, Ph.D., Co-Chair

Hai Liu, Ph.D., Co-Chair

# TABLE OF CONTENTS

LIST OF TABLES . . . . .	xi
LIST OF FIGURES . . . . .	xiv
Chapter 1 Introduction . . . . .	1
1.1 Time-varying exposure and multivariate longitudinal outcomes . . . . .	2
1.2 Flexible models for comparing cumulative effects of time-varying exposures . . . . .	3
1.3 A penalized Cox proportional hazards model with time-varying exposures and interaction term . . . . .	4
Chapter 2 Time-varying exposure and multivariate longitudinal outcomes . . . . .	6
2.1 Introduction . . . . .	6
2.2 A primary care depression screening cohort . . . . .	9
2.3 Methods . . . . .	12
2.3.1 A functional regression model for multivariate longitudinal out- comes . . . . .	12
2.3.2 Penalized spline . . . . .	15
2.3.3 Mixed model representation and parameter estimation . . . . .	17
2.3.4 Confidence interval of the coefficient function . . . . .	18
2.4 Simulation . . . . .	20
2.5 Data application . . . . .	31
2.6 Conclusion . . . . .	35
Chapter 3 Flexible models for comparing cumulative effects of time-dependent exposures . . . . .	37
3.1 Introduction . . . . .	37



3.2	A primary care depression screening cohort . . . . .	39
3.3	Method . . . . .	41
3.3.1	Cox's PH model with time-dependent WCE . . . . .	41
3.3.2	Estimation procedures . . . . .	44
3.3.3	Hypothesis testing . . . . .	45
3.3.4	Pointwise confidence band for the estimated weight function . . .	47
3.4	Simulation study . . . . .	47
3.4.1	Generating medication exposures and survival data . . . . .	47
3.4.2	Simulation results . . . . .	48
3.5	Data application . . . . .	55
3.6	Conclusion . . . . .	57
Chapter 4 A penalized Cox PH model with time-varying exposures and interaction		
	term . . . . .	60
4.1	Introduction . . . . .	60
4.2	A primary care depression screening cohort . . . . .	62
4.3	Methods . . . . .	66
4.3.1	A penalized Cox PH model with time-varying exposures and inter- action term . . . . .	66
4.3.2	Hypothesis test . . . . .	70
4.4	Simulation . . . . .	71
4.4.1	Simulation setup . . . . .	71
4.4.2	Simulation results . . . . .	73
4.5	Application . . . . .	91
4.6	Discussion . . . . .	97
Chapter 5 Conclusion . . . . .		99

BIBLIOGRAPHY . . . . .	102
CURRICULUM VITAE	

## LIST OF TABLES

2.1	Parameter estimates and estimated standard deviation over 100 replications under Scenario 1. Model (UR): Unconstrained coefficient function; Regression spline; Model (CR): Constrained coefficient function; Regression spline; Model (UP): Unconstrained coefficient function; Penalized spline; Model (CP): Constrained coefficient function; Penalized spline . . . . .	23
2.2	Parameter estimates and estimated standard deviation over 100 replications under Scenario 2. Model (UR): Unconstrained coefficient function; Regression spline; Model (CR): Constrained coefficient function; Regression spline; Model (UP): Unconstrained coefficient function; Penalized spline; Model (CP): Constrained coefficient function; Penalized spline . . . . .	28
2.3	Parameter estimates and estimated standard deviation over 100 replications under Scenario 3. Model (UR): Unconstrained coefficient function; Regression spline; Model (CR): Constrained coefficient function; Regression spline; Model (UP): Unconstrained coefficient function; Penalized spline; Model (CP): Constrained coefficient function; Penalized spline . . . . .	29
2.4	Parameter estimates and estimated standard deviation over 100 replications under Scenario 4. Model (UR): Unconstrained coefficient function; Regression spline; Model (CR): Constrained coefficient function; Regression spline; Model (UP): Unconstrained coefficient function; Penalized spline; Model (CP): Constrained coefficient function; Penalized spline . . . . .	29

2.5	Average mean squared error (AMSE) using the four estimation methods and average coverage rates (ACR) of 95% point-wise confidence bands estimated using the model with constrained coefficient function and penalized spline over 100 replications. Model (UR): Unconstrained coefficient function; Regression spline; Model (CR): Constrained coefficient function; Regression spline; Model (UP): Unconstrained coefficient function; Penalized spline; Model (CP): Constrained coefficient function; Penalized spline . . . . .	30
2.6	Parameter estimates for the bivariate models of longitudinal systolic and diastolic blood pressure measures with SSRIs use . . . . .	33
3.1	Estimated coefficients based on 1000 simulations with sample size 200 and 500	53
3.2	Hypothesis test of equal weight functions (sample size=500). Entries in the table are proportion of simulations where the null hypothesis is rejected based on 1000 simulations . . . . .	54
3.3	Model results using different medication exposure windows from 60 days up to 1920 days . . . . .	56
4.1	Average mean squared error (avgMSE) ( $\times 10^{-5}$ ) of coefficient functions and coefficient surface estimated by penalized Cox PH method over 200 replications	76
4.2	Empirical type I errors of Wald tests under sample size of 450 across 200 replications . . . . .	91
4.3	Comparison of demographic information among three groups. P-value is calculated using ANOVA for continuous variables and Chi-square test for binary variables . . . . .	93
4.4	Estimates and P-values of hypothesis test of no interaction effect under various pre-specified time interval . . . . .	94

4.5	Estimates and P-values of hypothesis test of no interaction effect under various	
	pre-specified time interval . . . . .	95

## LIST OF FIGURES

2.1	Time-varying medication exposure and repeated blood pressure measures from two randomly selected patients . . . . .	11
2.2	Four scenarios of coefficient functions used in simulations: $w^{(1)}(t)$ is indicated by solid lines; $w^{(2)}(t)$ is indicated by dashed lines. The coefficient functions were defined on reversed time axes so that day zero corresponds to the time of outcome measures and increasing days correspond to more distant past. Scenario 3 has identical coefficient functions for the bivariate outcomes. . .	22
2.3	Estimates of coefficient function under Scenario 1. White lines indicate true coefficient function. Top left: Unconstrained coefficient function using regression spline (Model UR); Top right: Constrained coefficient function using regression spline (Model CR); Bottom left: Unconstrained coefficient function using penalized spline (Model UP); Bottom right: Constrained coefficient function using penalized spline (Model CP) . . . . .	24
2.4	Estimates of coefficient function under Scenario 2. White lines indicate true coefficient function. Top left: Unconstrained coefficient function using regression spline (Model UR); Top right: Constrained coefficient function using regression spline (Model CR); Bottom left: Unconstrained coefficient function using penalized spline (Model UP); Bottom right: Constrained coefficient function using penalized spline (Model CP) . . . . .	25

2.5	Estimates of coefficient function under Scenario 3. White lines indicate true coefficient function. Top left: Unconstrained coefficient function using regression spline (Model UR); Top right: Constrained coefficient function using regression spline (Model CR); Bottom left: Unconstrained coefficient function using penalized spline (Model UP); Bottom right: Constrained coefficient function using penalized spline (Model CP) . . . . .	26
2.6	Estimates of coefficient function under Scenario 4. White lines indicate true coefficient function. Top left: Unconstrained coefficient function using regression spline (Model UR); Top right: Constrained coefficient function using regression spline (Model CR); Bottom left: Unconstrained coefficient function using penalized spline (Model UP); Bottom right: Constrained coefficient function using penalized spline (Model CP) . . . . .	27
2.7	Estimated coefficient functions with corresponding confidence bands under various exposure time windows . . . . .	34
2.8	Estimated coefficient functions with corresponding confidence bands under exposure time window of 360 Days with 8 interior knots . . . . .	35
3.1	Medication exposure of four randomly selected patients with event time (denoted by square) or censoring time (denoted by diamond) . . . . .	41
3.2	Six scenarios of true weight function for each medication. The time axis is the reversed time since exposure. The origin <i>Days</i> = 0 corresponds to current time, and increasing <i>Days</i> correspond to more distant past. All the weight functions are normalized. . . . .	49

3.3	A random sample of 100 normalized estimated weight functions (Sample size=500). Thick white lines indicate the true weight functions . Blue lines indicate weight functions for Medication A and red lines indicate weight functions for Medi- cation B. . . . .	51
3.4	Four randomly selected bootstrap confidence bands of inverted U weight func- tion (Scenario A4B4, Sample size=500): True weight function used to simu- lated data is denoted by solid line; Dotted line is the estimated weight function from original simulated data set; Dashed lines are upper or lower bound of 95% bootstrap confidence band . . . . .	52
3.5	Estimated weight functions with 240-day window for antidepressant SSRIs and TCAs user as well as 95 percent bootstrap confidence band: solid lines indicate estimated weight functions and dashed lines indicate 95% confidence bands (SSRIs in blue and TCAs in red); Dashed horizontal line is the zero line. . . . .	57
4.1	Examples of medication exposure from four typical patients. Upper left: pa- tient who took TCA exclusively; Upper right: patient who take SSRI exclu- sively; Lower left: patient who switched from TCA to SSRI; Lower right: patient who took TCA and SSRI sequentially . . . . .	65
4.2	Four scenarios of coefficient functions used in simulations: $w_{(A)}(t)$ is indicated by solid lines; $w_{(B)}(t)$ is indicated by dashed lines. The coefficient functions were defined on reversed time axes so that day zero corresponds to the time of outcome measures and increasing days correspond to more distant past. Scenario 3 has identical coefficient functions. . . . .	72



4.3	Two 3D surfaces with associated contour plots used to generate the interaction surface with a pre-specified time window of 30 days. Surfaces were defined on reversed time axes, where x-axis indicates the exposure time for medication A and y-axis represents medication B. Upper surface has constant highest values on the diagonal. Middle surface has decreasing values with increasing days. Lower surface is a weighted combination of upper surface and middle surface.	74
4.4	Estimated coefficient functions and mean of estimated coefficient surface in Scenario 1 with sample size of 300. True coefficient functions are denoted as white solid lines. . . . .	77
4.5	Estimated coefficient functions and mean of estimated coefficient surface in Scenario 1 with sample size of 450. True coefficient functions are denoted as white solid lines. . . . .	78
4.6	Estimated coefficient functions and mean of estimated coefficient surface in Scenario 1 with sample size of 600. True coefficient functions are denoted as white solid lines. . . . .	79
4.7	Estimated coefficient functions and mean of estimated coefficient surface in Scenario 2 with sample size of 300. True coefficient functions are denoted as white solid lines. . . . .	80
4.8	Estimated coefficient functions and mean of estimated coefficient surface in Scenario 2 with sample size of 450. True coefficient functions are denoted as white solid lines. . . . .	81
4.9	Estimated coefficient functions and mean of estimated coefficient surface in Scenario 2 with sample size of 600. True coefficient functions are denoted as white solid lines. . . . .	82

4.10	Estimated coefficient functions and mean of estimated coefficient surface in Scenario 3 with sample size of 300. True coefficient functions are denoted as white solid lines. . . . .	83
4.11	Estimated coefficient functions and mean of estimated coefficient surface in Scenario 3 with sample size of 450. True coefficient functions are denoted as white solid lines. . . . .	84
4.12	Estimated coefficient functions and mean of estimated coefficient surface in Scenario 3 with sample size of 600. True coefficient functions are denoted as white solid lines. . . . .	85
4.13	Estimated coefficient functions and mean of estimated coefficient surface in Scenario 4 with sample size of 300. True coefficient functions are denoted as white solid lines. . . . .	86
4.14	Estimated coefficient functions and mean of estimated coefficient surface in Scenario 4 with sample size of 450. True coefficient functions are denoted as white solid lines. . . . .	87
4.15	Estimated coefficient functions and mean of estimated coefficient surface in Scenario 4 with sample size of 600. True coefficient functions are denoted as white solid lines. . . . .	88
4.16	Sample point-wise confidence bands for Scenario 1 and Scenario 2. Red dashed lines are 2.5% and 97.5% point-wise quantiles. Blue dashed lines are mean estimates. The true coefficient functions are denoted as black solid lines. . .	89
4.17	Sample point-wise confidence bands for Scenario 3 and Scenario 4. Red dashed lines are 2.5% and 97.5% point-wise quantiles. Blue dashed lines are mean estimates. The true coefficient functions are denoted as black solid lines. . .	90

4.18	Estimates of coefficient functions for both TCA and SSRI under various pre-specified time interval. Black lines indicate TCA. Red lines indicate SSRI .	96
------	---	----

# Chapter 1

## Introduction

An electronic medical record (EMR) is a systematic collection of electronic health information about an individual patient. It is a record in digital format that is theoretically capable of being shared across different health care settings, including enormous quantities of clinical data such as medical diagnosis, laboratory testing and medication dispensing information. They have been increasingly used in many health systems around the country, which offers an unprecedented research opportunity for monitoring disease development, progression and treatment. In particular, medication dispensing data allow researchers to explore complex relationships among long-term medication use, disease progression and potential side-effects in large patient populations. However, these data also pose challenges to existing statistical models because both medication exposure status and its intensity vary over time, known as time-varying exposure in many studies. Many medical journals published studies used summarized exposure measures such as current dose (Smitten et al., 2008), average daily dose (Wolfe et al., 2006), or a simple sum of past exposures (Stranges et al., 2006). Few studies on long-term medication effects fully utilize the medication information contained in EMR data.

We have developed several novel approaches for assessing the effect of time-varying exposures on longitudinal and survival outcomes. First, a penalized functional regression model was developed to estimate the effect of time-varying exposure on multivariate longitudinal outcomes. Second, for survival outcomes, a regression spline based model was proposed in the Cox proportional hazards (PH) framework to compare disease risk among different types of time-varying exposures. Finally, a penalized spline based Cox PH model with functional

interaction terms was developed to estimate potential interaction effect between multiple medication classes.

### **1.1 Time-varying exposure and multivariate longitudinal outcomes**

To accommodate time-varying exposures statistically, Breslow et al. (1983) and Thomas (1988) were among the first ones to propose a weighted cumulative exposure (WCE) measure with predetermined weights to summarize time-varying exposures. Later on, various approaches have also been proposed using nonparametric methods to estimate the weight function in generalized linear models (Berhane et al., 2008; Hauptmann et al., 2000) and Cox proportional hazards models (Gasparrini, 2013; Sylvestre and Abrahamowicz, 2009). In the meanwhile, functional regression models have been developed for time-varying covariates where the functional coefficients can be interpreted as the weight functions for time-varying exposures. Zhang et al. (2007) evaluated the effect of time-varying follicle stimulating hormone during a menstrual cycle on total hip bone mineral density and used a two-stage functional mixed model with periodic cubic smoothing splines to estimate the coefficient function. Schipper et al. (2008) proposed a generalized monotonic functional mixed model (GMFMM) with regression splines to relate a functional measure of radiation dose to a normal tissue complication outcome. A Bayesian GMFMM was also proposed for a modified weight function (Schipper et al., 2007). Bhadra et al. (2012) presented a Bayesian semi-parametric approach for time-varying exposure on a binary outcome in a case-control study. For time-varying exposure and a single longitudinal outcome, Goldsmith et al. (2012) developed a penalized functional regression model and compared a likelihood based method and a Bayesian method for model estimation and inference.

In modeling medication effects, it is of scientific interest to determine the effect of long-term medication use on a number of correlated longitudinal outcomes such as repeated

systolic and diastolic blood pressure measures. In such situations, separate modeling of systolic and diastolic outcomes may not be appropriate, as the two outcomes are biologically correlated and mutually influential (Guo and Carlin, 2004; Liu et al., 2012). However, few studies discussed the development of a multivariate longitudinal functional regression model to estimate the exposure effects while controlling for the interdependency across correlated outcomes.

In Chapter 2, we propose a multivariate semiparametric model on longitudinally measured outcomes with constrained coefficient functions for time-varying exposures.. In particular, we use penalized splines to flexibly estimate the coefficient functions and shared random effects to model the correlations among the multivariate longitudinal outcomes. We show that the proposed model has a mixed effect model representation so that the model estimation and inferences can be achieved using standard statistical softwares.

## **1.2 Flexible models for comparing cumulative effects of time-varying exposures**

Many extensions have been made in defining weighted cumulative exposure recently. For example, Berhane et al. (2008) proposed a tensor product spline model to jointly model age, latency and exposure response effects for protracted time-dependent occupational exposure histories in the Colorado plateau uranium miners cohort. Gasparrini (2013) noted out that equivalent approaches were previously established in time series analysis. Analogous to the weight function, the definition of a distributed lag function was introduced in distributed lag models (DLMs) and distributed lag nonlinear models (DLNMs) in time series analysis.

Another existing problem that has not been addressed thus far is that in clinical practice patients often receive different types of medications intended for the same medical condition and it is of interest to compare the effects of these medication exposures on relevant health

outcomes. However, little work has been done in comparing the effects of multiple medication exposures over an extended period using time to event outcomes. In Chapter 3, we extend the regression spline-based method in Sylvestre and Abrahamowicz (2009) to model WCE in multiple exposures settings where the weight functions for each exposure type can be estimated, where a direct comparison between multiple exposures can be considered in this unified modeling framework.

### **1.3 A penalized Cox proportional hazards model with time-varying exposures and interaction term**

One limitation of the methodology discussed in Chapter 3 is that we did not consider patients who were on multiple types of medication simultaneously intended for the same medical condition, which is quite common in clinical practice. Generally, patients take multiple drugs during their follow-up periods. In clinical practice, it is usual for patients to receive different types of medications intended for the same medical condition. For example, depressed patients may switch from one type of antidepressants to another if the treatment fails to show satisfactory efficacy. Patients with hypertension may take more than one class of antihypertensive medications at the same time. Therefore, it is of interest to assess the effects of multiple medications and their potential interaction effects on relevant health outcomes.

Many functional regression models included only one functional covariate, such as in Zhang et al. (2007), Schipper et al. (2008) and Bhadra et al. (2012). Goldsmith et al. (2012) and Ferraty and Vieu (2009) included two or more functional covariates through additive models. However, not much work has been done to study the interaction effect between functional covariates. When the outcome of interest is a scalar, Fuchs et al. (2015) proposed a penalized scalar-on-functions regression with interaction term, which extends the model

with only main effects in Wood (2011). To the best of our knowledge, the estimation of interaction effects between functional covariates has not received much attention in survival analysis setting.

In Chapter 4, we proposed a penalized Cox PH model with functional covariates and their interaction. In particular, the coefficient functions for functional covariates have been expanded using cubic B-spline basis for the main effects and tensor product splines for the interaction effect. Penalized partial likelihood method was used to estimate the coefficient functions and surfaces nonparametrically. Hypothesis tests of no interaction or no effect of time-varying exposure were also discussed, based on Wald's method.



## Chapter 2

### Time-varying exposure and multivariate longitudinal outcomes

#### 2.1 Introduction

Electronic medical records (EMR) capture enormous quantities of clinical data including medical diagnosis, laboratory testing, medication dispensing information and they have been increasingly used in many health systems around the country. The availability of EMR data offers an unprecedented research opportunity for monitoring disease development, progression and treatment. In particular, medication dispensing data allow researchers to explore complex relationships among long-term medication use, disease progression and potential side-effects in large patient populations. Since many clinical trials for medication approval were conducted in restrictive patient populations due to stringent exclusion criteria with limited numbers of outcomes and over a relatively short period of time, the long-term effect of medications on many health outcomes in the general patient population can be better studied using EMR data.

Electronic medication dispensing data typically contain information on medication names, dosage, and length for the dispensed medication. However, few studies on long-term medication effects fully utilize the information contained in EMR data. Many published studies used summarized exposure measures such as current dose (Smitten et al., 2008), average daily dose (Wolfe et al., 2006), or a simple sum of past exposures (Stranges et al., 2006). Breslow et al. (1983) and Thomas (1988) were among the first ones to propose a weighted cumulative exposure (WCE) measure with predetermined weights to summarize time-varying exposures. Abrahamowicz et al. (2006) used a parametric WCE framework within the Cox's proportional hazards (PH) model to study the association between a time-varying exposure

and event risk. Applications using WCE by simple parametric functions can also be found in works by VACEK (1997) , Langholz et al. (1999), and Richardson (2009). Various approaches have also been proposed using nonparametric methods to estimate the weight function in generalized linear models (Berhane et al., 2008; Hauptmann et al., 2000) and Cox proportional hazards models (Gasparrini, 2013; Sylvestre and Abrahamowicz, 2009).

For scalar outcomes, functional regression models have been developed for time-varying covariates where the functional coefficients can be interpreted as the weight functions for time-varying exposures. Zhang et al. (2007) evaluated the effect of time-varying follicle stimulating hormone during a menstrual cycle on total hip bone mineral density and used a two-stage functional mixed model with periodic cubic smoothing splines to estimate the coefficient function. Schipper et al. (2008) proposed a generalized monotonic functional mixed model (GMFMM) with regression splines to relate a functional measure of radiation dose to a normal tissue complication outcome. A Bayesian GMFMM was also proposed for a modified weight function (Schipper et al., 2007). Bhadra et al. (2012) presented a Bayesian semi-parametric approach for time-varying exposure on a binary outcome in a case-control studies. For time-varying exposure and a single longitudinal outcome, Goldsmith et al. (2012) developed a penalized functional regression model and compared a likelihood based method and a Bayesian method for model estimation and inference.

In modeling medication effects, it is often of interest to determine the effect of long-term medication use on a number of correlated longitudinal outcomes such as repeated systolic and diastolic blood pressure measures. In such situations, separate modeling of systolic and diastolic outcomes may not be appropriate, as the two outcomes are biologically correlated and mutually influential (Guo and Carlin, 2004; Liu et al., 2012). Therefore, a multivariate longitudinal functional regression model needs to be developed in order to

estimate the exposure effects more appropriately while controlling for the interdependency across correlated outcomes.

In addition, when modeling the effect of exposure to medication or environmental toxins, there is often prior evidence that exposures occurring long before an outcome may have negligible influence, which suggests that the coefficient function should be bounded toward zero at a distant boundary. By imposing such constraint would also avoid potential boundary issues in estimating the coefficient functions using regression splines as encountered in other studies (Sylvestre and Abrahamowicz, 2009). Schipper et al. (2008) considered similar constraints on the coefficient function which force the functional value to be zero at zero dose.

In this Chapter, we consider modeling multivariate longitudinal outcomes for the estimation of time-varying exposures with constrained coefficient functions. In particular, we use penalized splines to flexibly estimate the coefficient functions and shared random effects to model the correlations among the multivariate longitudinal outcomes. We show that the proposed model has a mixed effect model representation so that the model estimation and inferences can be achieved using standard statistical softwares.

This Chapter is organized as follows. Section 2 describes data from a primary care depression study as a motivating example for our methods. Section 3 introduces the constrained multivariate functional regression model and discusses parameter estimation and point-wise confidence bands for the estimated coefficient functions. Section 4 presents the results from a simulation study. In Section 5, we apply the proposed methods to the real data and estimate the association between antidepressant use and longitudinal blood pressure measures. We conclude the paper with a discussion in Section 6.

## 2.2 A primary care depression screening cohort

A primary care patient cohort was assembled in 1991 as part of a depression screening study in primary care clinics at Wishard Health Service. From 1991 to 1993, patients aged 60 years or older in Wishard Health Service were consented for depression screening during their regular clinical visits to their primary care physicians. A total of 4,413 primary care patients were initially contacted, of whom 115 refused; 57 were not eligible due to severe cognitive impairment; 284 were not eligible because they were non-English speaking, in prison, in a nursing home, or had a hearing impairment; 3,957 patients were enrolled in the study. Details of the study have been published elsewhere (Callahan et al., 1994; McDonald et al., 1999).

For enrolled patients, information including diagnosis of medical conditions, blood pressure measures, laboratory test measures and medications order and dispensing was collected by the Regenstrief Medical Record System (RMRS) (McDonald et al., 1999). The RMRS is one of the first electronic medical record systems in the country and has been actively used for research purposes. Originally implemented in 1973 and used continuously since then, the RMRS serves as the day-to-day electronic medical record system at Wishard Hospital and its community clinics - the third largest safety-net health care system in the U.S. Medication dispensing information includes medication name, daily dose, and duration for each dispensed medication. Our interest in this paper is to examine the association between a class of antidepressants and longitudinally measured blood pressure levels.

Antidepressant is one of the most commonly prescribed medication groups in the United States (Lindsley, 2012). An older class of antidepressants, tricyclic antidepressants (TCAs), have been shown to have detrimental effect on cardiovascular function by inhibiting cardiovascular  $\text{Na}(+)$ ,  $\text{Ca}(2+)$  and  $\text{K}(+)$  channels often leading to life-threatening arrhythmia (Glassman, 1984; Jefferson, 1975). A newer class of antidepressants, selective serotonin

reuptake inhibitors (SSRIs), became the preferred treatment for depression due to its comparable efficacy with TCAs and its superior tolerability. SSRIs were hypothesized to show a different cardiovascular effect from TCAs due to their pharmacologic profile (Bergstrom et al., 1988). However, there have been reports of first-degree atrioventricular block, prolonged QTc interval, and orthostatic hypotension in SSRI-treated patients suggesting that SSRIs may also have important cardiac and vascular effects (de la Torre et al., 2001; Pacher and Kecskemeti, 2004). Given that many of these studies were conducted in the laboratory setting with brief SSRIs treatment, EMR data with detailed medication dispensing information offer a unique opportunity to examine the effect of long-term SSRIs use on cardiac functions in an elderly patient population.

A total of 159 patients had SSRIs dispensing records and blood pressure measures in the EMR. In order to capture the effect of SSRIs on newly diagnosed depression patients, we excluded patients with a history of depression at enrollment resulting in a total of 110 patients on SSRIs. To derive comparable dosages among different SSRI medications, we first standardized medication doses using daily dose divided by the recommended minimum dose for each medication (Damush et al., 2008). Medication dispensing data on antidepressants displays varied patterns for both dose levels and the duration of medication use. In Figure 2.1, medication dosage and duration as well as repeated blood pressure measures from two randomly selected patients in the analysis sample are displayed. Longitudinally measured blood pressure levels from one year before the first SSRI dispensing to the end of patient's follow-up were included in our analyses. The number of blood pressure measures per patient ranges from 1 to 104 with a median of 18.

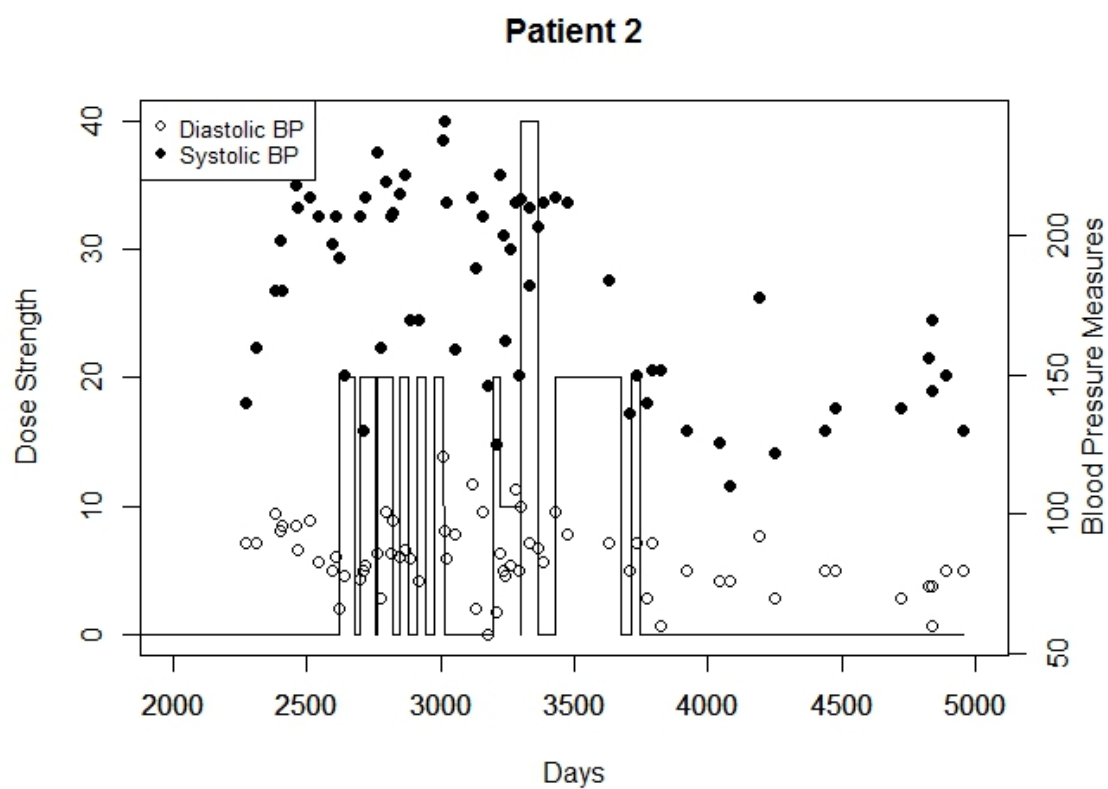
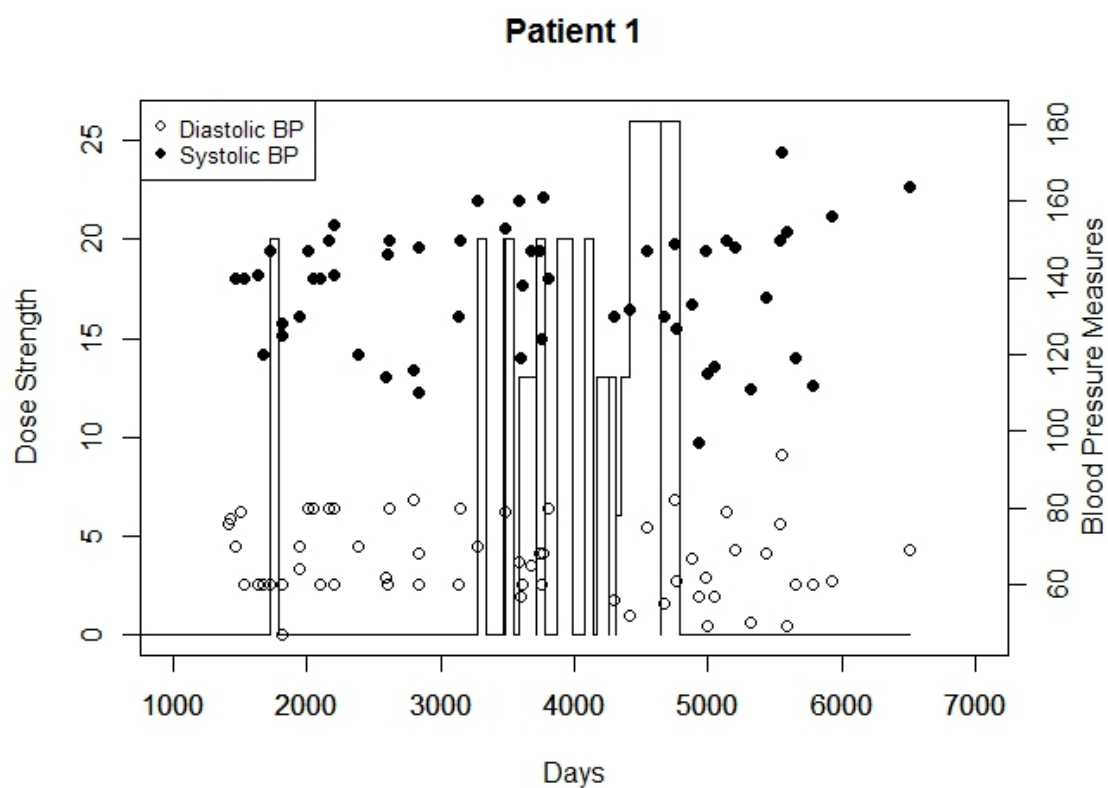


Figure 2.1: Time-varying medication exposure and repeated blood pressure measures from two randomly selected patients

## 2.3 Methods

### 2.3.1 A functional regression model for multivariate longitudinal outcomes

Let  $Y^{(k)}(u_{ij}), k = 1, \dots, K$  be the  $k$ th outcome from the  $i$ th subject of the  $j$ th visit measured at time  $u_{ij}$ , where  $i = 1, \dots, m$  and  $j = 1, \dots, n_i$ .  $D_i(t)$  is the time-varying exposure for the  $i$ th subject at time  $t$  and  $\mathbf{X}_{ij}$  is a covariate vector for the  $i$ th subject. We propose the following functional regression model for multivariate longitudinal outcomes:

$$\begin{cases} Y^{(1)}(u_{ij}) = \int_0^{u_{ij}} D_i(u_{ij} - t)w^{(1)}(t)dt + \mathbf{X}_{ij}^T \boldsymbol{\beta}^{(1)} + b_i^{(1)} + \varepsilon_{ij}^{(1)} \\ \dots \\ Y^{(K)}(u_{ij}) = \int_0^{u_{ij}} D_i(u_{ij} - t)w^{(K)}(t)dt + \mathbf{X}_{ij}^T \boldsymbol{\beta}^{(K)} + b_i^{(K)} + \varepsilon_{ij}^{(K)} \end{cases} \quad (2.1)$$

where  $w^{(1)}(t), \dots, w^{(K)}(t)$  are unknown coefficient functions,  $\boldsymbol{\beta}^{(1)}, \dots, \boldsymbol{\beta}^{(K)}$  are parameters of fixed effects for the covariates  $\mathbf{X}_{ij}$ ; Denote  $\mathbf{b}_i = (b_i^{(1)}, \dots, b_i^{(K)})^T$  as the random subject effects which are used to account for both temporal correlations among the repeated measurements from the same subject, and the correlation among the multiple outcomes. We assume that the subject-specific random effects follow multivariate normal distribution, that is,  $\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_b)$  and

$$\boldsymbol{\Sigma}_b = \begin{pmatrix} \sigma_b^{(11)} & \dots & \sigma_b^{(1K)} \\ \dots & \dots & \dots \\ \sigma_b^{(1K)} & \dots & \sigma_b^{(KK)} \end{pmatrix}$$

where  $\sigma_b^{(kl)}$  is the covariance between the random subject effects for outcomes  $Y^{(k)}(u_{ij})$  and  $Y^{(l)}(u_{ij})$  from the same subject measured across all visits.  $\varepsilon_{ij}^{(1)}, \dots, \varepsilon_{ij}^{(K)}$  are random errors following independent normal distribution  $(\varepsilon_{ij}^{(1)}, \dots, \varepsilon_{ij}^{(K)})^T \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$ , where  $\boldsymbol{\Sigma}_\varepsilon = \text{diag}(\sigma_{\varepsilon 1}^2, \dots, \sigma_{\varepsilon K}^2)$ .

Note that the coefficient function was defined on a reversed time axis, so the origin zero corresponds to the time when the outcome was measured. Thus  $w^{(k)}(0)$  can be interpreted as the effect of one unit of exposure strength at the concurrent time of the  $k$ th outcome measure and  $w^{(k)}(t)$  the effect of one unit exposure strength at  $t$  time units prior to the measurement of the  $k$ th outcome. In this model framework, cumulative effects for each outcome over  $u_{ij}$  length are modeled as

$$\int_0^{u_{ij}} D_i(u_{ij} - t)w^{(k)}(t)dt \quad (k = 1, \dots, K)$$

.

In many existing applications of exposure assessment, there was plenty of evidence for diminished effects from distant exposures, which can be linked to the compound's half life. Therefore, we assume that past exposure occurring  $a$  time units before the outcome measures has a negligible effect on current outcomes. This assumption is equivalent to requiring that the coefficient functions will smoothly go to zero at the distant boundary of the exposure interval. One natural way to impose such a constraint is to define the coefficient function  $w(t)$  as an integral:  $w(t) = \int_{-a}^{-t} s(c)dc, t \in [0, a]$ , where  $s(c) = \sum v_l B_l(c)$  and  $B_l(c)$  are some splines defined by using  $L$  interior knots equally spaced on  $[-a, 0]$ , and  $v_l$ 's are the coefficients for the spline bases. Without loss of generality, in this study we adopt the truncated cubic spline bases:  $1, c, c^2, c^3, (c - \kappa_1)_+^3, \dots, (c - \kappa_L)_+^3$ . Therefore,  $w(t)$  is the coefficient function defined on the interval  $[0, a]$  with  $w(a) = 0$  at a distant boundary  $a$  time units away from the time of the outcome measure,  $t$ .



It follows that model (2.1) can be rewritten as

$$\begin{cases} Y^{(1)}(u_{ij}) = \mathbf{M}_{ij}^T \mathbf{v}^{(1)} + \mathbf{X}_{ij}^T \boldsymbol{\beta}^{(1)} + b_i^{(1)} + \varepsilon_{ij}^{(1)} \\ \dots \\ Y^{(K)}(u_{ij}) = \mathbf{M}_{ij}^T \mathbf{v}^{(K)} + \mathbf{X}_{ij}^T \boldsymbol{\beta}^{(K)} + b_i^{(K)} + \varepsilon_{ij}^{(K)} \end{cases} \quad (2.2)$$

where

$$\mathbf{M}_{ij} = (M_{ij1}, \dots, M_{ij(L+4)})^T$$

and

$$M_{ijl} = \int_0^{u_{ij}} D_i(u_{ij} - t) \int_{-a}^{-t} B_l(c) dc dt$$

Let  $\mathbf{Y} = ((\mathbf{Y}^{(1)})^T, \dots, (\mathbf{Y}^{(K)})^T)^T$  be the response variable vectors, where  $\mathbf{Y}^{(k)} = (Y_{1,1}^{(k)}, \dots, Y_{m,n_m}^{(k)})^T, k = 1, \dots, K$ . Let  $\mathbf{b} = ((\mathbf{b}^{(1)})^T, \dots, (\mathbf{b}^{(K)})^T)^T$  be the vectors of subject-specific random effects, where  $\mathbf{b}^{(k)} = (b_1^{(k)}, \dots, b_m^{(k)})^T, k = 1, \dots, K$ . We denote  $\boldsymbol{\varepsilon}^{(k)} = (\varepsilon_{1,1}^{(k)}, \dots, \varepsilon_{m,n_m}^{(k)})^T, k = 1, \dots, K$ , and let  $\boldsymbol{\varepsilon} = ((\boldsymbol{\varepsilon}^{(1)})^T, \dots, (\boldsymbol{\varepsilon}^{(K)})^T)^T$  be the random error vectors.

We denote the design matrix with parameter vector  $\mathbf{v} = ((\mathbf{v}^{(1)})^T, \dots, (\mathbf{v}^{(K)})^T)^T$ , where  $\mathbf{v}^{(k)} = (v_1^{(k)}, \dots, v_{L+4}^{(k)})^T$  in model (2.2) as  $\tilde{\mathbf{M}} = \mathbf{I}_K \otimes \mathbf{M}$ , where  $\mathbf{I}_n$  is the identity matrix of dimension  $n$ ,  $\otimes$  denotes the Kronecker product and  $\mathbf{M} = [\mathbf{M}_{ij}^T]_{1 \leq j \leq n_i; 1 \leq i \leq m}$ . Similarly, we set up the design matrix of fixed effect as  $\tilde{\mathbf{X}} = \mathbf{I}_K \otimes \mathbf{X}$  with corresponding parameter vector  $\boldsymbol{\beta} = ((\boldsymbol{\beta}^{(1)})^T, \dots, (\boldsymbol{\beta}^{(K)})^T)^T$ , where  $\mathbf{X} = [\mathbf{X}_{ij}^T]_{1 \leq j \leq n_i; 1 \leq i \leq m}$ . The design matrix of random effect is  $\tilde{\mathbf{Z}} = \mathbf{I}_K \otimes \mathbf{Z}_b$  so that the elements of  $\mathbf{Z}_b \mathbf{b}_k$  corresponding to subject  $i$  are equal to  $b_i^{(k)}$ . Therefore the functional regression model for multivariate longitudinal

outcomes can be written in matrix form as:

$$\mathbf{Y} = \tilde{\mathbf{M}}\mathbf{v} + \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\mathbf{Z}}\mathbf{b} + \boldsymbol{\varepsilon} \quad (2.3)$$

where  $\tilde{\mathbf{M}}$  and  $\tilde{\mathbf{X}}$  are the design matrices of the fixed effects  $\mathbf{v}$  and  $\boldsymbol{\beta}$ , respectively.  $\mathbf{b} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_b \otimes \mathbf{I}_m)$  is the random effects vector, and random errors are  $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon \otimes \mathbf{I}_N)$ , where  $N = \sum_{i=1}^m n_i$  is the total number of observations.

### 2.3.2 Penalized spline

Model (2.3) can be regarded as a mixed model with  $\mathbf{v}$  and  $\boldsymbol{\beta}$  as fixed effects, and  $\mathbf{b}$  as random effects. The conditional distribution of the outcome variables given the random effects,  $\mathbf{b}$ , is normal as

$$\mathbf{Y}|\mathbf{b} \sim N(\tilde{\mathbf{M}}\mathbf{v} + \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\mathbf{Z}}\mathbf{b}, \boldsymbol{\Sigma}_\varepsilon \otimes \mathbf{I}_N)$$

where  $\mathbf{b} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_b \otimes \mathbf{I}_m)$ . The best linear unbiased predictors (BLUP) of  $\mathbf{v}$ ,  $\boldsymbol{\beta}$  and  $\mathbf{b}$  can be obtained by minimizing:

$$(\mathbf{Y} - \tilde{\mathbf{M}}\mathbf{v} - \tilde{\mathbf{X}}\boldsymbol{\beta} - \tilde{\mathbf{Z}}\mathbf{b})^T \mathbf{R}^{-1} (\mathbf{Y} - \tilde{\mathbf{M}}\mathbf{v} - \tilde{\mathbf{X}}\boldsymbol{\beta} - \tilde{\mathbf{Z}}\mathbf{b}) + \mathbf{b}^T \mathbf{G}^{-1} \mathbf{b}$$

where  $\mathbf{R} = \boldsymbol{\Sigma}_\varepsilon \otimes \mathbf{I}_N$  and  $\mathbf{G} = \boldsymbol{\Sigma}_b \otimes \mathbf{I}_m$  (Robinson, 1991).

The flexibility of the coefficient function can be controlled by the number of interior knots once the spline order is fixed. A method used to avoid overfitting by the regression splines is to impose a constraint on the coefficient of higher order basis functions as  $\sum_{l=5}^{L+4} (v_l^{(k)})^2 < C^{(k)}, k = 1, \dots, K$  with a pre-chosen constant  $C^{(k)}$  resulting in a penalized spline estimate. Using Lagrange multiplier, imposing the penalty on  $\mathbf{v}_r^{(k)} = (v_5^{(k)}, \dots, v_{L+4}^{(k)})^T$  is equivalent to

minimizing

$$(\mathbf{Y} - \tilde{\mathbf{M}}\mathbf{v} - \tilde{\mathbf{X}}\boldsymbol{\beta} - \tilde{\mathbf{Z}}\mathbf{b})^T \mathbf{R}^{-1}(\mathbf{Y} - \tilde{\mathbf{M}}\mathbf{v} - \tilde{\mathbf{X}}\boldsymbol{\beta} - \tilde{\mathbf{Z}}\mathbf{b}) + \mathbf{b}^T \mathbf{G}^{-1} \mathbf{b} + \sum_{k=1}^K \lambda_k^2 (\mathbf{v}_r^{(k)})^T \mathbf{v}_r^{(k)} \quad (2.4)$$

for certain tuning parameter  $\lambda_k \geq 0$ . Compared to regression splines without the constraint, penalized splines are not as sensitive to the number of knots and location of the knots as long as the knots are adequately spaced and the number of knots is sufficiently large (Hastie and Tibshirani, 1990; Ruppert, 2002).

The tuning parameters  $\lambda_k$  can be chosen by cross-validation (Linhart and Zucchini, 1986). Alternatively, if we assume that  $\mathbf{v}_r^{(k)}$  as a set of random variables follows a normal distribution,  $\mathbf{v}_r^{(k)} \sim N(\mathbf{0}, \sigma_{v^{(k)}}^2 \mathbf{I}_L)$ , the BLUP estimates of  $\mathbf{v}, \boldsymbol{\beta}$  and  $\mathbf{b}$  can be obtained by minimizing

$$(\mathbf{Y} - \tilde{\mathbf{M}}\mathbf{v} - \tilde{\mathbf{X}}\boldsymbol{\beta} - \tilde{\mathbf{Z}}\mathbf{b})^T \mathbf{R}^{-1}(\mathbf{Y} - \tilde{\mathbf{M}}\mathbf{v} - \tilde{\mathbf{X}}\boldsymbol{\beta} - \tilde{\mathbf{Z}}\mathbf{b}) + \mathbf{b}^T \mathbf{G}^{-1} \mathbf{b} + \sum_{k=1}^K \frac{1}{\sigma_{v^{(k)}}^2} (\mathbf{v}_r^{(k)})^T \mathbf{v}_r^{(k)} \quad (2.5)$$

which is identical to (2.4) when  $\lambda_k^2 = \frac{1}{\sigma_{v^{(k)}}^2}$ . Therefore, by treating  $\mathbf{v}_r^{(k)}$  as random effects from a normal distribution, we obtain the estimate of the tuning parameter with penalized spline using maximum likelihood instead of cross-validation. Compared with the cross-validation approach for choosing tuning parameters through penalized estimation procedure, ML-based methods offer computational advantageous (Kohn et al., 1991). Moreover, this representation of the penalized spline as a BLUP in a mixed model allows the use of mixed model estimation implemented in standard statistical software.

### 2.3.3 Mixed model representation and parameter estimation

The proposed functional regression model in Section 2.3.2 using penalized spline for multivariate longitudinal outcomes can be written as a mixed effects model as follows:

$$\mathbf{Y} = \tilde{\mathbf{M}}_f \mathbf{v}_f + \tilde{\mathbf{X}} \boldsymbol{\beta} + \tilde{\mathbf{M}}_r \mathbf{v}_r + \tilde{\mathbf{Z}} \mathbf{b} + \boldsymbol{\varepsilon} \quad (2.6)$$

where  $\mathbf{v}_f = (v_1^{(1)}, \dots, v_4^{(1)}, \dots, v_1^{(K)}, \dots, v_4^{(K)})^T$  is the coefficient vector of lower order bases estimated as fixed effects.  $\mathbf{v}_r = ((\mathbf{v}_r^{(1)})^T, \dots, (\mathbf{v}_r^{(K)})^T)^T$  is the coefficient vector of higher order bases and can be treated as random effects as  $\mathbf{b}$ , together with variance-covariance matrix:

$$\text{Cov} \begin{bmatrix} \mathbf{v}_r^{(1)} \\ \dots \\ \mathbf{v}_r^{(K)} \\ \mathbf{b} \\ \boldsymbol{\varepsilon} \end{bmatrix} = \begin{bmatrix} \sigma_{v^{(1)}}^2 \mathbf{I}_L & \mathbf{0} & 0 & 0 & 0 \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 0 & \mathbf{0} & \sigma_{v^{(K)}}^2 \mathbf{I}_L & 0 & 0 \\ 0 & \mathbf{0} & 0 & \mathbf{G} & 0 \\ 0 & \mathbf{0} & 0 & 0 & \mathbf{R} \end{bmatrix}$$

where  $\mathbf{G}$  and  $\mathbf{R}$  are defined previously.  $\mathbf{G}$  measures the between-subject variation,  $\mathbf{R}$  measures the within-subject variation and  $\sigma_{v^k}^2$  controls the smoothness of the  $k$ th coefficient function.

In practice the time-varying exposure  $D_i, i = 1, \dots, m$  are usually measured at discrete time points, e.g. daily medication intake. Therefore,  $M_{ijl} = \int_0^{u_{ij}} D_i(u_{ij} - t) \int_{-a}^{-t} B_l(c) dc dt$  in model (2.2) can be approximated by  $M_{ijl} = \sum_0^{u_{ij}} D_i(u_{ij} - t) (\sum_{-a}^{-t} B_l(c) \Delta c) \Delta t$ , where  $\Delta c$  and  $\Delta t$  are discrete time units for the coefficient function and time-varying exposures

respectively. Substituting the discrete time notations into model (2.2), we obtain

$$\left\{ \begin{array}{l} Y^{(1)}(u_{ij}) = \sum_{l=1}^{L+4} v_l^{(1)} [\sum_0^{u_{ij}} D_i(u_{ij} - t) (\sum_{-a}^{-t} B_l(c) \Delta c) \Delta t] + \mathbf{X}_{ij}^T \boldsymbol{\beta}^{(1)} + b_i^{(1)} + \varepsilon_{ij}^{(1)} \\ \dots \\ Y^{(K)}(u_{ij}) = \sum_{l=1}^{L+4} v_l^{(K)} [\sum_0^{u_{ij}} D_i(u_{ij} - t) (\sum_{-a}^{-t} B_l(c) \Delta c) \Delta t] + \mathbf{X}_{ij}^T \boldsymbol{\beta}^{(K)} + b_i^{(K)} + \varepsilon_{ij}^{(K)} \end{array} \right. \quad (2.7)$$

.

#### 2.3.4 Confidence interval of the coefficient function

Our primary interest is to estimate the coefficient functions  $w^{(1)}(t)$  to  $w^{(K)}(t)$ . Based on model (2.7), for a given time  $t_0$ ,  $w^{(k)}(t_0)$ ,  $k = 1, \dots, K$  is estimated as

$$\hat{w}^{(k)}(t_0) = (\mathbf{C}(t_0))^T \hat{\mathbf{v}}^{(k)} = (\mathbf{C}_{\mathbf{f}}(t_0))^T \hat{\mathbf{v}}_{\mathbf{f}}^{(k)} + (\mathbf{C}_{\mathbf{r}}(t_0))^T \hat{\mathbf{v}}_{\mathbf{r}}^{(k)} \quad (2.8)$$

where

$$\mathbf{C}(t_0) = ((\mathbf{C}_{\mathbf{f}}(t_0))^T, (\mathbf{C}_{\mathbf{r}}(t_0))^T)^T$$

$$\mathbf{C}_{\mathbf{f}}(t_0) = (C_1(t_0), \dots, C_4(t_0))^T$$

$$\mathbf{C}_{\mathbf{r}}(t_0) = (C_5(t_0), \dots, C_{L+4}(t_0))^T$$

and

$$C_l(t_0) = \sum_{c=-a}^{-t_0} B_l(c) \Delta c$$

is a cumulative summation of bases  $B_l$  up to time  $t_0$ .  $\hat{\mathbf{v}}^{(k)}$  includes predicted values of random effects  $\hat{\mathbf{v}}_{\mathbf{r}}^{(k)}$  and estimated values of fixed effects  $\hat{\mathbf{v}}_{\mathbf{f}}^{(k)}$ .

Based on the large sample theory in mixed models (Ruppert et al., 2003), the limiting distribution of  $\hat{w}^{(k)}$  evaluated at  $t_0$  is approximately normal

$$\frac{\hat{w}^{(k)}(t_0) - w^{(k)}(t_0)}{\sqrt{\mathbf{v}\hat{\mathbf{a}}\mathbf{r}(\hat{w}^{(k)}(t_0) - w^{(k)}(t_0))}} \sim N(0, 1)$$

therefore, an approximate  $100(1 - \alpha)\%$  confidence interval for  $w(t_0)$  can be constructed as  $\hat{w}^{(k)}(t_0) \pm z(1 - \frac{\alpha}{2})\sqrt{\mathbf{v}\hat{\mathbf{a}}\mathbf{r}(\hat{w}^{(k)}(t_0) - w^{(k)}(t_0))}$  where

$$\mathbf{v}\hat{\mathbf{a}}\mathbf{r}(\hat{w}^{(k)}(t_0) - w^{(k)}(t_0)) = (\mathbf{C}(t_0))^T \mathbf{Cov} \begin{pmatrix} \hat{v}_f^{(k)} - v_f^{(k)} \\ \hat{v}_r^{(k)} - v_r^{(k)} \end{pmatrix} \mathbf{C}(t_0) \quad (2.9)$$

The variance-covariance matrix in equation (2.9) is a sub-matrix of the following variance-covariance matrix (Searle et al., 2009):

$$\mathbf{Cov} \begin{pmatrix} \hat{v}_f - v_f \\ \hat{\beta} - \beta \\ \hat{v}_r - v_r \\ \hat{b} - b \end{pmatrix} = \mathbf{Cov} \begin{pmatrix} \hat{v}_f \\ \hat{\beta} \\ \hat{v}_r - v_r \\ \hat{b} - b \end{pmatrix}$$

with

$$\mathbf{Var} \begin{pmatrix} \hat{v}_f - v_f \\ \hat{\beta} - \beta \end{pmatrix} = ([\tilde{\mathbf{M}}_f \ \tilde{\mathbf{X}}]^T \mathbf{V}^{-1} [\tilde{\mathbf{M}}_f \ \tilde{\mathbf{X}}])^{-1}$$

,

$$\mathbf{Var} \begin{pmatrix} \hat{v}_r - v_r \\ \hat{b} - b \end{pmatrix} = \mathbf{D} - \mathbf{D}[\tilde{\mathbf{M}}_r \ \tilde{\mathbf{Z}}]^T \mathbf{P}[\tilde{\mathbf{M}}_r \ \tilde{\mathbf{Z}}] \mathbf{D}$$

and

$$\text{Cov} \left( \begin{pmatrix} \hat{v}_f - v_f \\ \hat{\beta} - \beta \end{pmatrix}, \begin{pmatrix} \hat{v}_r - v_r \\ \hat{b} - b \end{pmatrix}^T \right) = -([\tilde{\mathbf{M}}_f \ \tilde{\mathbf{X}}]^T \mathbf{V}^{-1} [\tilde{\mathbf{M}}_f \ \tilde{\mathbf{X}}])^{-1} [\tilde{\mathbf{M}}_f \ \tilde{\mathbf{X}}]^T \mathbf{V}^{-1} [\tilde{\mathbf{M}}_r \ \tilde{\mathbf{Z}}] \mathbf{D}$$

where

$$\mathbf{D} = \begin{pmatrix} \sigma_{v(1)}^2 \mathbf{I}_L & \mathbf{0} & 0 & 0 \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ 0 & \mathbf{0} & \sigma_{v(\kappa)}^2 \mathbf{I}_L & 0 \\ 0 & \mathbf{0} & 0 & \mathbf{G} \end{pmatrix}, \mathbf{V} = [\tilde{\mathbf{M}}_r \ \tilde{\mathbf{Z}}] \mathbf{D} [\tilde{\mathbf{M}}_r \ \tilde{\mathbf{Z}}]^T + \mathbf{R}$$

and

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1} [\tilde{\mathbf{M}}_f \ \tilde{\mathbf{X}}] ([\tilde{\mathbf{M}}_f \ \tilde{\mathbf{X}}]^T \mathbf{V}^{-1} [\tilde{\mathbf{M}}_f \ \tilde{\mathbf{X}}])^{-1} [\tilde{\mathbf{M}}_f \ \tilde{\mathbf{X}}]^T \mathbf{V}^{-1}$$

## 2.4 Simulation

To evaluate the performance of our proposed functional regression model for multivariate longitudinal outcomes, simulation studies were conducted. Without loss of generality, we considered bivariate outcomes. The timing of the exposure measured by days with treatment or interruption in the unit of 7 days, was generated from a lognormal distribution with mean 0.5 and standard deviation 0.8 on the log scale, i.e.  $\log(\text{duration}) \sim N(0.5, 0.8^2)$ , and rounded up to the nearest integer. The strengths of the time-varying exposure  $D_i$  were assumed to be constant over each treatment period (in 7 day unit), which could take values of 0.5, 1, 1.5, 2, 2.5, or 3 during the follow-up period with equal probabilities. Two correlated outcome  $(Y^{(1)}(u_{ij}), Y^{(2)}(u_{ij}))^T$  were generated from the following model for sample size of

50 ( $i = 1, \dots, 50$ ) with random number of observations per subject of mean 10 ( $j = 1, \dots, n_i$ ):

$$\begin{cases} Y^{(1)}(u_{ij}) = \alpha_1 + \int_0^{u_{ij}} D_i(u_{ij} - t)w^{(1)}(t)dt + b_i^{(1)} + \varepsilon_{ij}^{(1)} \\ Y^{(2)}(u_{ij}) = \alpha_2 + \int_0^{u_{ij}} D_i(u_{ij} - t)w^{(2)}(t)dt + b_i^{(2)} + \varepsilon_{ij}^{(2)} \end{cases} \quad (2.10)$$

where intercepts are  $\alpha_1 = 72$  and  $\alpha_2 = 122$ . Random variables  $(b_{ij}^{(1)}, b_{ij}^{(2)})^T \sim N(\mathbf{0}, \mathbf{\Sigma}_b)$  and  $(\varepsilon_{ij}^{(1)}, \varepsilon_{ij}^{(2)})^T \sim N(\mathbf{0}, \mathbf{\Sigma}_\varepsilon)$  have variance covariance matrix as

$$\mathbf{\Sigma}_b = \begin{pmatrix} 50 & 10 \\ 10 & 50 \end{pmatrix}, \mathbf{\Sigma}_\varepsilon = \begin{pmatrix} 100 & 0 \\ 0 & 400 \end{pmatrix}$$

Four coefficient functions with a pre-specified time window of 30 days were considered (Figure 4.2), which include zero function ( $w(t) = 0, t \in [0, 30]$ ) indicating no exposure effect over the time window, decreasing function ( $w(t) = \sin(\frac{\pi}{2}\frac{t}{30} + \frac{\pi}{2}), t \in [0, 30]$ ) indicating greater effect closer to event time with less effect for distant exposures, and U-shape function ( $w(t) = \sin(\frac{3\pi}{2}\frac{t}{30} + \frac{\pi}{2}), t \in [0, 30]$ ) suggesting greater effect for recent exposures and decreasing effects for distant exposures. All coefficient functions smoothly go to zero at the distant boundary of the exposure interval. We generated 100 data sets under each scenario as shown in Figure 4.2.

To further evaluate our proposed method in Section 2.3.2, we compared four models: 1. Unconstrained coefficient function using regression spline (Model UR); 2. Constrained coefficient function using regression spline (Model CR); 3. Unconstrained coefficient function using penalized spline (Model UP); 4. Constrained coefficient function using penalized spline (Model CP). Six equally spaced interior knots were used for all models.

Parameter estimates and sample standard deviation of estimates over 100 replications are provided in Table 2.1 to Table 2.4. Models with unconstrained coefficient function have



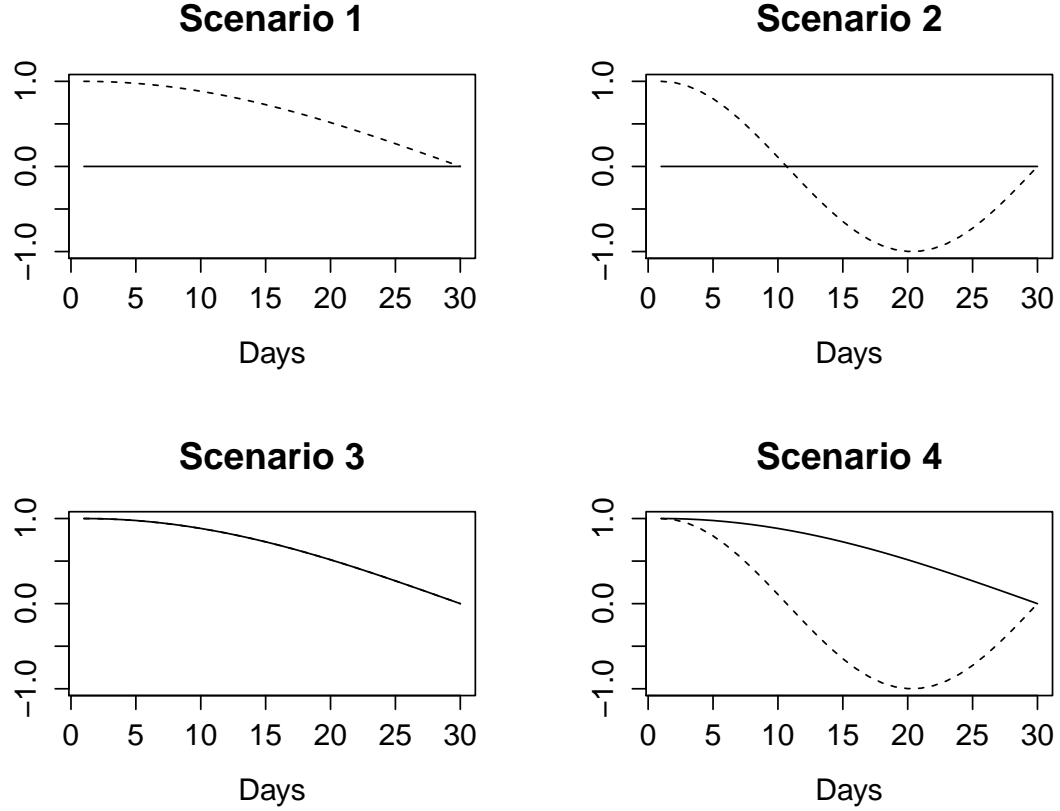


Figure 2.2: Four scenarios of coefficient functions used in simulations:  $w^{(1)}(t)$  is indicated by solid lines;  $w^{(2)}(t)$  is indicated by dashed lines. The coefficient functions were defined on reversed time axes so that day zero corresponds to the time of outcome measures and increasing days correspond to more distant past. Scenario 3 has identical coefficient functions for the bivariate outcomes.

biased estimates of  $\sigma_\varepsilon^{(1)}$  and have large sample standard deviations for  $\sigma_\varepsilon^{(1)}$  and  $\sigma_b^{(12)}$ , the covariance between  $b^{(1)}$  and  $b^{(2)}$  in Model (2.10).

Table 2.1: Parameter estimates and estimated standard deviation over 100 replications under Scenario 1. Model (UR): Unconstrained coefficient function; Regression spline; Model (CR): Constrained coefficient function; Regression spline; Model (UP): Unconstrained coefficient function; Penalized spline; Model (CP): Constrained coefficient function; Penalized spline

Model		$\alpha_1$	$\alpha_2$	$\sqrt{\sigma_b^{(11)}}$	$\sqrt{\sigma_b^{(22)}}$	$\sigma_b^{(12)}$	$\sigma_\varepsilon^{(1)}$	$\sigma_\varepsilon^{(2)}$
	True	72	122	$\sqrt{50}$	$\sqrt{50}$	10	10	20
CP	Mean	72.085	122.016	6.984	6.709	8.436	9.917	19.984
	SD	1.128	1.802	0.824	1.361	1.162	0.291	0.692
UP	Mean	72.063	122.063	6.984	6.721	10.184	7.121	19.983
	SD	1.145	1.779	0.854	1.355	9.471	6.943	0.688
CR	Mean	72.025	121.978	7.203	6.202	8.971	9.935	20.033
	SD	1.165	1.871	0.930	2.091	0.944	0.289	0.796
UR	Mean	72.060	122.047	7.013	6.768	10.616	7.114	19.977
	SD	1.146	1.794	0.873	1.376	11.874	6.937	0.691

Figure 2.3 to Figure 2.6 shows 100 estimated coefficient functions obtained from the four models under four scenarios. In each of the panels, the true coefficient function used to generate data was in white. The estimated coefficient functions generally captures the shape of the true functions. Instability at the tail was observed in models with unconstrained coefficient function regardless of the types of splines. Constraining the coefficient function seems to have reduced the variation at the tail considerably.

Mean squared errors (MSE) of the estimated coefficient functions  $\hat{w}(t)$  were calculated as  $MSE = \frac{1}{30} \sum_{t=0}^{t=30} (\hat{w}(t) - w(t))^2$ . Table 4.2 displays, under each scenario, average mean squared error (AMSE) for the four models. The CP model consistently yielded one of the lowest AMSE values across all cases. Table 4.2 also displays average coverage rates (ACR) of 95% point-wise confidence bands estimated for Model CP. In most cases, ACRs for model CP are close to the nominal level of 0.95.

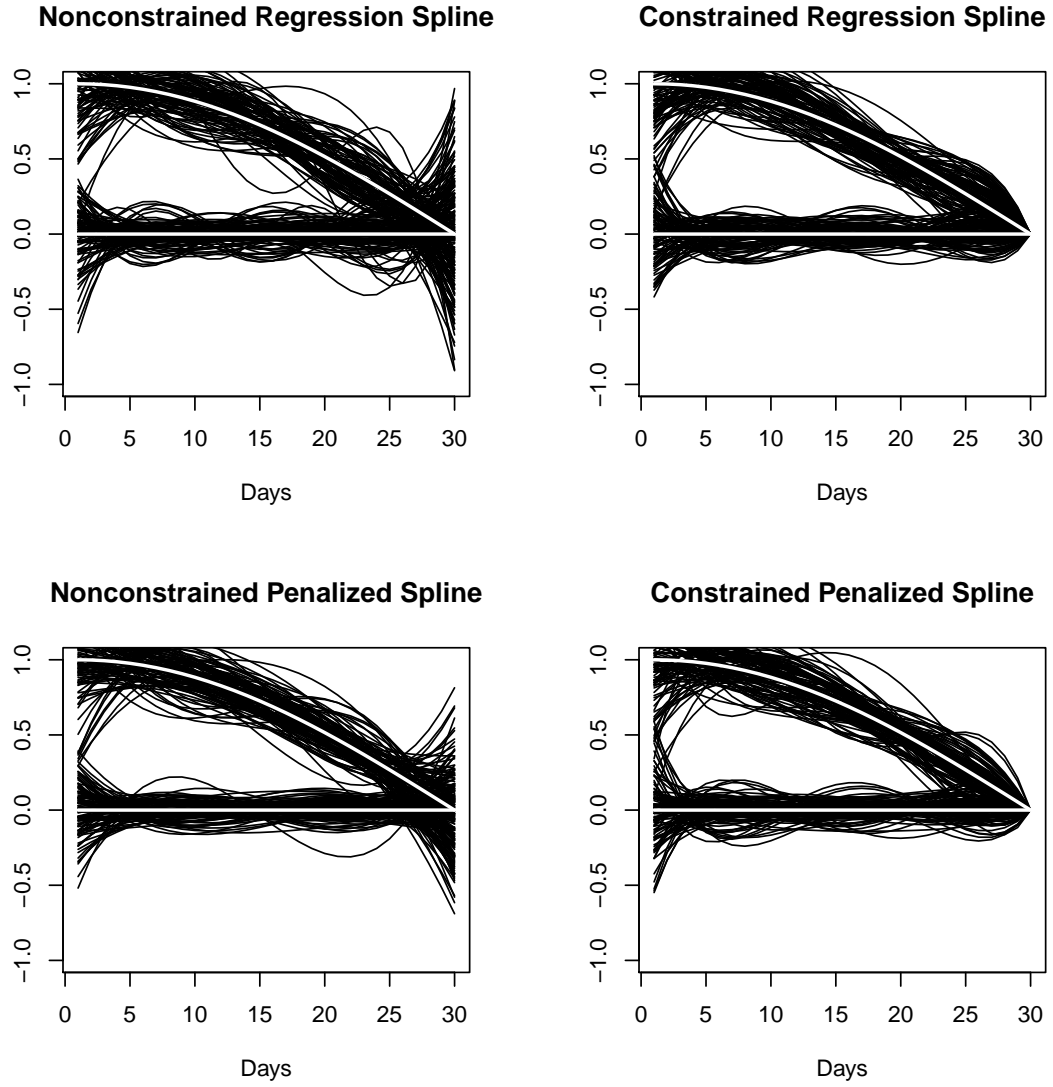


Figure 2.3: Estimates of coefficient function under Scenario 1. White lines indicate true coefficient function. Top left: Unconstrained coefficient function using regression spline (Model UR); Top right: Constrained coefficient function using regression spline (Model CR); Bottom left: Unconstrained coefficient function using penalized spline (Model UP); Bottom right: Constrained coefficient function using penalized spline (Model CP)

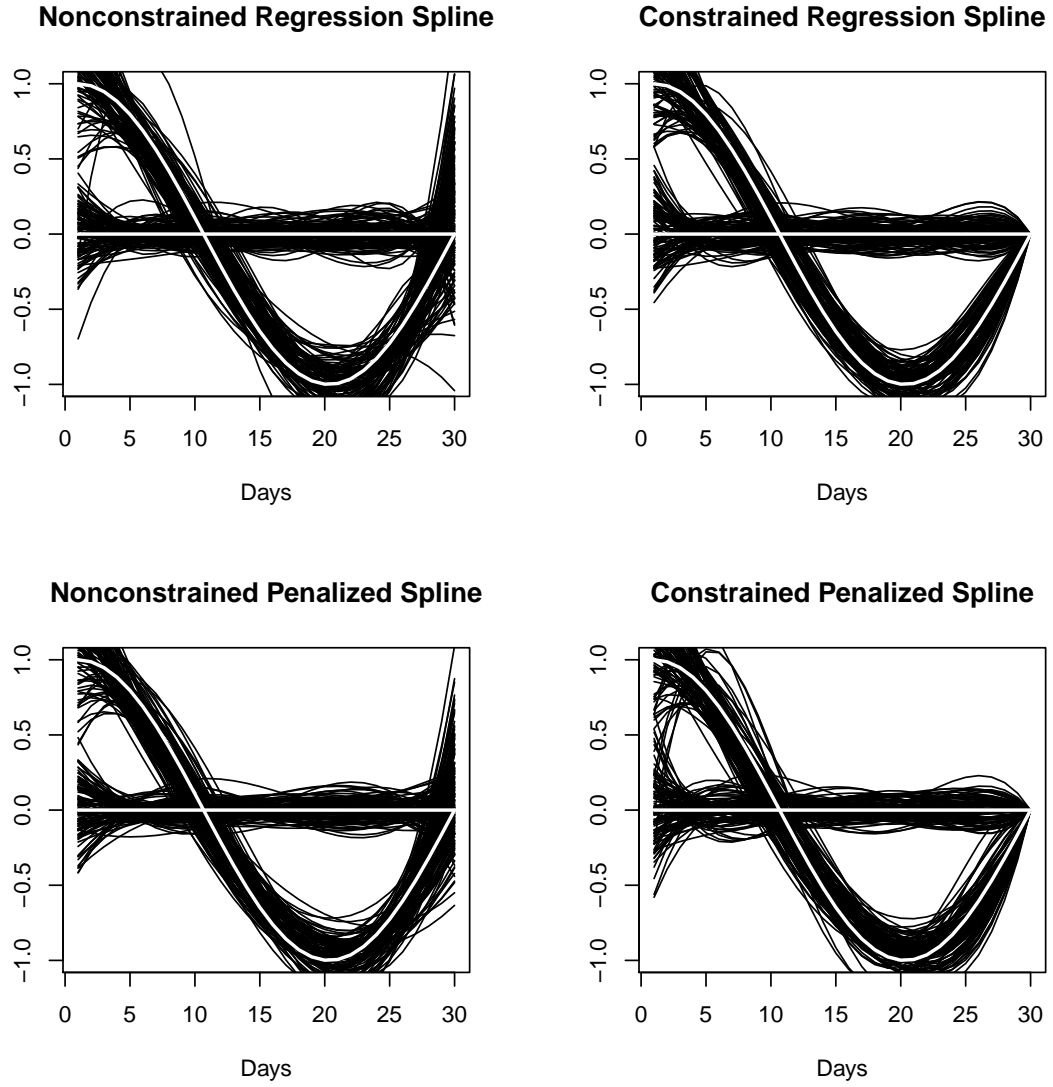


Figure 2.4: Estimates of coefficient function under Scenario 2. White lines indicate true coefficient function. Top left: Unconstrained coefficient function using regression spline (Model UR); Top right: Constrained coefficient function using regression spline (Model CR); Bottom left: Unconstrained coefficient function using penalized spline (Model UP); Bottom right: Constrained coefficient function using penalized spline (Model CP)

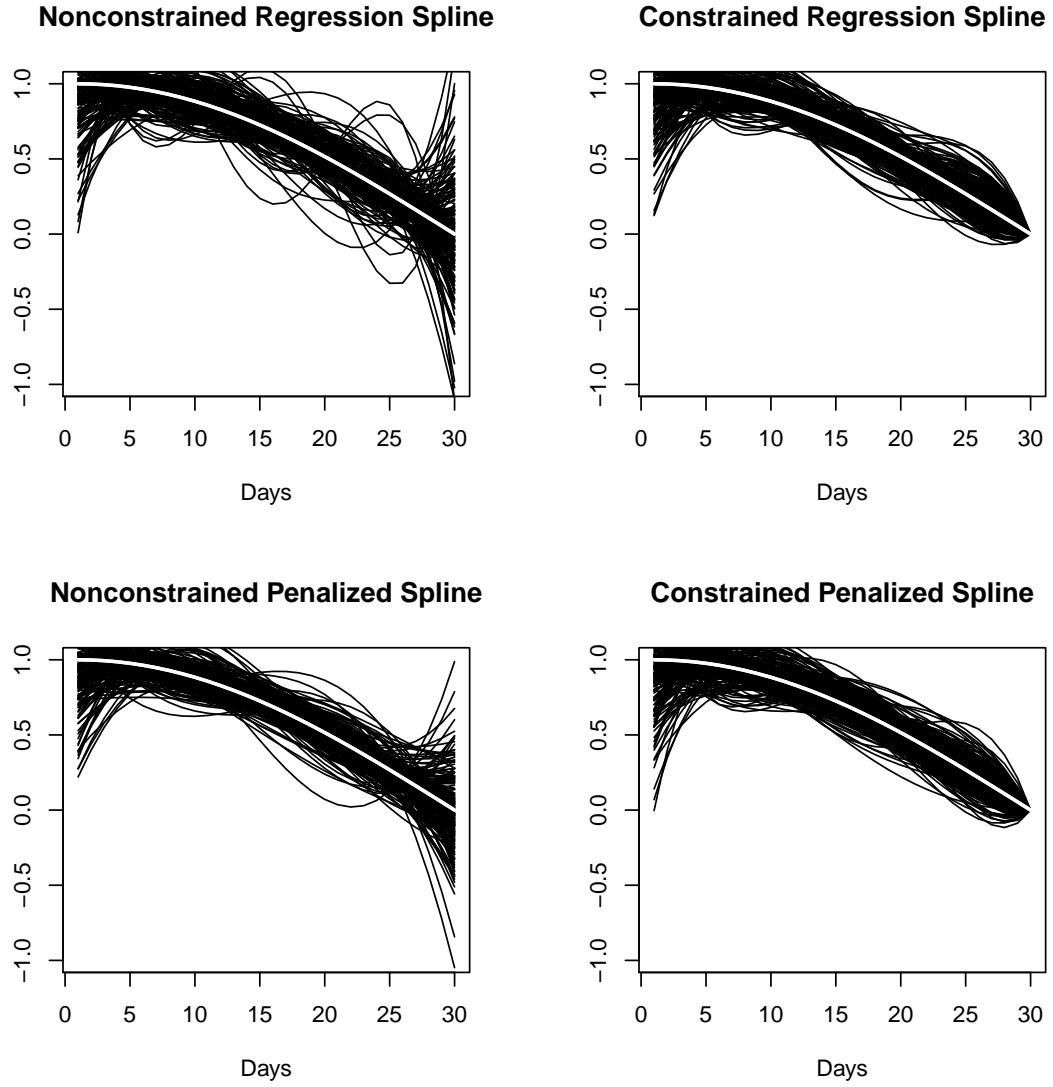


Figure 2.5: Estimates of coefficient function under Scenario 3. White lines indicate true coefficient function. Top left: Unconstrained coefficient function using regression spline (Model UR); Top right: Constrained coefficient function using regression spline (Model CR); Bottom left: Unconstrained coefficient function using penalized spline (Model UP); Bottom right: Constrained coefficient function using penalized spline (Model CP)

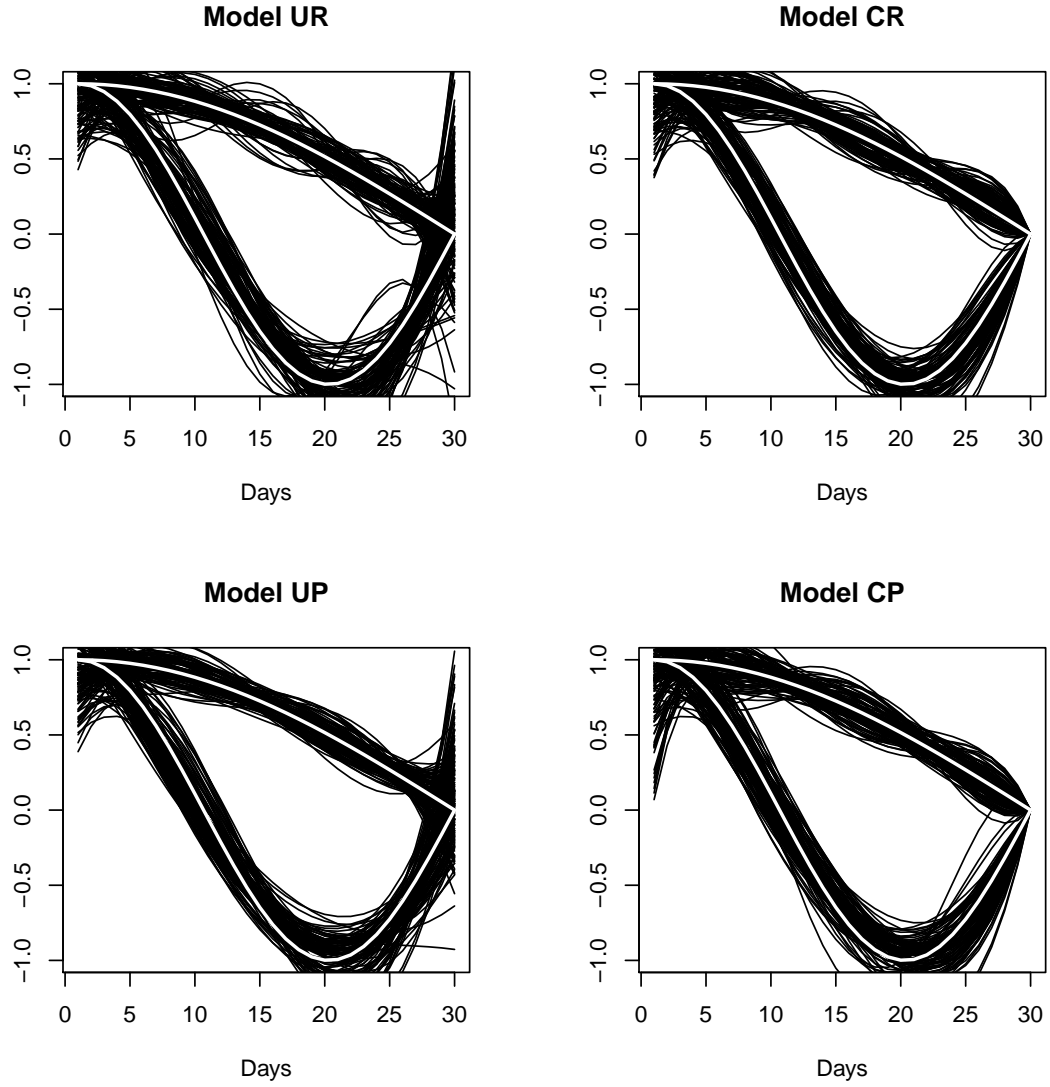


Figure 2.6: Estimates of coefficient function under Scenario 4. White lines indicate true coefficient function. Top left: Unconstrained coefficient function using regression spline (Model UR); Top right: Constrained coefficient function using regression spline (Model CR); Bottom left: Unconstrained coefficient function using penalized spline (Model UP); Bottom right: Constrained coefficient function using penalized spline (Model CP)

Table 2.2: Parameter estimates and estimated standard deviation over 100 replications under Scenario 2. Model (UR): Unconstrained coefficient function; Regression spline; Model (CR): Constrained coefficient function; Regression spline; Model (UP): Unconstrained coefficient function; Penalized spline; Model (CP): Constrained coefficient function; Penalized spline

Model		$\alpha_1$	$\alpha_2$	$\sqrt{\sigma_b^{(11)}}$	$\sqrt{\sigma_b^{(22)}}$	$\sigma_b^{(12)}$	$\sigma_\varepsilon^{(1)}$	$\sigma_\varepsilon^{(2)}$
True		72	122	$\sqrt{50}$	$\sqrt{50}$	10	10	20
CP	Mean	72.087	122.183	6.849	6.862	8.244	9.936	19.897
	SD	1.340	1.561	0.952	1.154	1.903	0.303	0.532
UP	Mean	72.107	122.043	6.859	6.887	8.485	5.943	19.888
	SD	1.357	1.565	0.943	1.205	10.540	8.008	0.532
CR	Mean	72.092	122.240	7.015	6.113	8.936	9.945	19.904
	SD	1.309	1.659	1.028	2.883	0.707	0.304	0.670
UR	Mean	72.100	122.029	6.860	6.898	8.477	5.939	19.882
	SD	1.356	1.555	0.930	1.199	10.789	8.005	0.534

The simulation results indicate that our proposed method provided adequate parameter estimates, were able to capture the shapes of the true coefficient functions, had low MSE and provided satisfactory coverage rates.

Table 2.3: Parameter estimates and estimated standard deviation over 100 replications under Scenario 3. Model (UR): Unconstrained coefficient function; Regression spline; Model (CR): Constrained coefficient function; Regression spline; Model (UP): Unconstrained coefficient function; Penalized spline; Model (CP): Constrained coefficient function; Penalized spline

Model		$\alpha_1$	$\alpha_2$	$\sqrt{\sigma_b^{(11)}}$	$\sqrt{\sigma_b^{(22)}}$	$\sigma_b^{(12)}$	$\sigma_\varepsilon^{(1)}$	$\sigma_\varepsilon^{(2)}$
	True	72	122	$\sqrt{50}$	$\sqrt{50}$	10	10	20
CP	Mean	71.917	121.986	6.924	6.934	8.353	9.991	19.880
	SD	1.312	1.754	0.915	1.309	1.279	0.321	0.626
UP	Mean	71.914	121.981	6.940	6.937	9.554	7.590	19.872
	SD	1.342	1.770	0.923	1.295	7.878	6.536	0.624
CR	Mean	71.897	121.906	7.096	6.627	8.961	10.006	19.835
	SD	1.336	1.810	0.950	1.627	0.718	0.325	0.634
UR	Mean	71.906	121.987	6.961	6.971	9.840	7.585	19.864
	SD	1.343	1.777	0.951	1.319	9.509	6.532	0.625

Table 2.4: Parameter estimates and estimated standard deviation over 100 replications under Scenario 4. Model (UR): Unconstrained coefficient function; Regression spline; Model (CR): Constrained coefficient function; Regression spline; Model (UP): Unconstrained coefficient function; Penalized spline; Model (CP): Constrained coefficient function; Penalized spline

Model		$\alpha_1$	$\alpha_2$	$\sqrt{\sigma_b^{(11)}}$	$\sqrt{\sigma_b^{(22)}}$	$\sigma_b^{(12)}$	$\sigma_\varepsilon^{(1)}$	$\sigma_\varepsilon^{(2)}$
True		72	122	$\sqrt{50}$	$\sqrt{50}$	10	10	20
CP	Mean	72.229	122.140	6.810	7.011	8.333	9.943	19.898
	SD	1.323	1.857	0.842	1.274	1.638	0.357	0.583
UP	Mean	72.194	121.954	6.801	6.967	8.007	5.987	19.894
	SD	1.366	1.856	0.845	1.293	9.799	7.985	0.584
CR	Mean	72.173	122.042	6.974	6.552	8.861	9.954	19.867
	SD	1.366	1.908	0.929	2.014	0.841	0.359	0.647
UR	Mean	72.177	121.995	6.806	6.962	7.385	5.983	19.886
	SD	1.359	1.823	0.837	1.289	10.809	7.978	0.581



Table 2.5: Average mean squared error (AMSE) using the four estimation methods and average coverage rates (ACR) of 95% point-wise confidence bands estimated using the model with constrained coefficient function and penalized spline over 100 replications. Model (UR): Unconstrained coefficient function; Regression spline; Model (CR): Constrained coefficient function; Regression spline; Model (UP): Unconstrained coefficient function; Penalized spline; Model (CP): Constrained coefficient function; Penalized spline

Scenario	Coefficient function	AMSE				ACR
		(UR)	(CR)	(UP)	(CP)	(CP)
Scenario 1	$w^{(1)}(t)$	0.010	0.006	0.007	0.006	0.950
	$w^{(2)}(t)$	0.027	0.015	0.018	0.014	0.953
Scenario 2	$w^{(1)}(t)$	0.008	0.006	0.007	0.006	0.954
	$w^{(2)}(t)$	0.035	0.015	0.028	0.021	0.935
Scenario 3	$w^{(1)}(t)$	0.009	0.008	0.007	0.008	0.909
	$w^{(2)}(t)$	0.035	0.017	0.022	0.017	0.952
Scenario 4	$w^{(1)}(t)$	0.010	0.006	0.007	0.006	0.950
	$w^{(2)}(t)$	0.027	0.015	0.018	0.014	0.953

## 2.5 Data application

In this section we revisit the motivating data set we described in Section 2.2. We used the following model to examine the effects of antidepressants on diastolic and systolic blood pressure:

$$\begin{cases} DBP(u_{ij}) = \mathbf{X}_{ij}^T \beta^{(D)} + \int_0^{u_{ij}} D_i(u_{ij} - t) w^{(D)}(t) dt + b_i^{(D)} + \varepsilon_{ij}^{(D)} \\ SBP(u_{ij}) = \mathbf{X}_{ij}^T \beta^{(S)} + \int_0^{u_{ij}} D_i(u_{ij} - t) w^{(S)}(t) dt + b_i^{(S)} + \varepsilon_{ij}^{(S)} \end{cases} \quad (2.11)$$

where  $DBP(u_{ij})$  and  $SBP(u_{ij})$  represent diastolic and systolic blood pressure measures for  $i$ th subject at time  $u_{ij}$ .  $\mathbf{X}_{ij}^T$  represent covariates including intercept, time, gender and the use of antihypertensive.  $D_i(t)$  is the time-varying SSRIs exposure.  $w^{(D)}(t)$  and  $w^{(S)}(t)$  are the coefficient functions for diastolic and systolic blood pressure measures respectively.  $b_i^{(D)}$  and  $b_i^{(S)}$  are correlated random subject effects.  $\varepsilon_{ij}^{(D)}$  and  $\varepsilon_{ij}^{(S)}$  are random errors. Repeated blood pressure measures from one year before the first SSRIs dispensing to one year after the last SSRIs dispensing were included in the analysis, resulting in a total number of 1714 measures for each blood pressure type.

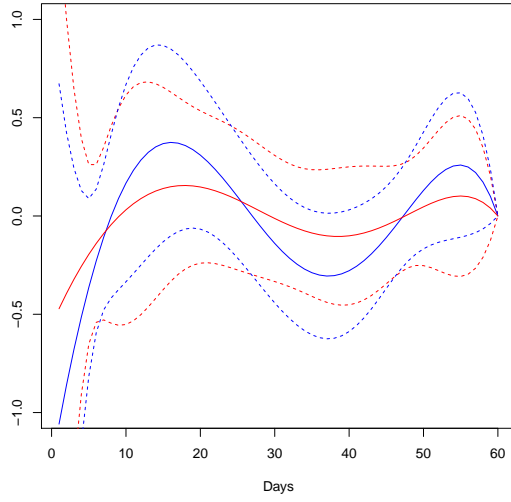
We estimated the coefficient functions under various pre-specified exposure time windows using truncated cubic bases with 6 equally spaced interior knots. Parameter estimates of fixed and random effects and variance components are included in Table 4.4. The estimates are relatively stable across different time windows. Since these models with different time intervals have the same number of parameters, the best fitting model can be selected based on the likelihood. The model with a 60-day exposure window has the smallest negative likelihood and is selected to be the best fitting model (Table 4.4). Subject effects of systolic blood pressure have greater variability ( $\hat{\sigma}_b^{(S)} = 20.48(SD = 1.23)$ ) than diastolic ( $\hat{\sigma}_b^{(D)} = 10.44(SD = 0.65)$ ) with a covariance of 8.58. The random error associated with systolic

blood pressure also has larger variability ( $\hat{\sigma}_\varepsilon^{(S)} = 20.03(SD = 0.36)$ ) than diastolic blood pressure ( $\hat{\sigma}_\varepsilon^{(D)} = 12.15(SD = 0.22)$ ).

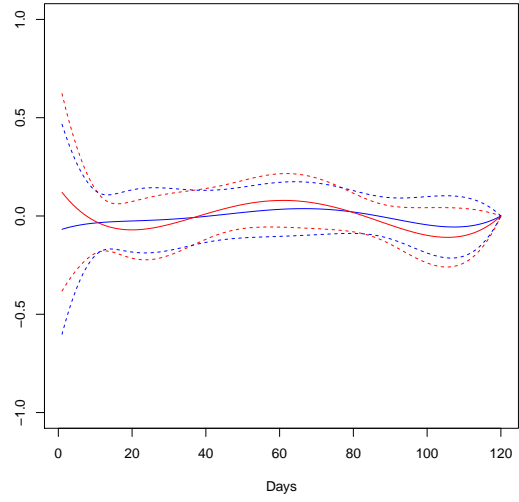
Figure 2.7 shows estimated coefficient functions with confidence bands under various exposure windows. It can be seen that all confidence bands contain the zero line, which indicates that there is no significant SSRIs effect on diastolic or either systolic blood pressures. We observe that estimated coefficients show less fluctuations under longer exposure time window. When we increased the number of interior knots to 8 under the 360-day exposure time window, similar shape of the estimated coefficient functions were observed as compared with using 6 interior knots 2.8. Therefore, the finding of no significant SSRIs effect on BP is not likely to be sensitive to the choice of the number of knots.

Table 2.6: Parameter estimates for the bivariate models of longitudinal systolic and diastolic blood pressure measures with SSRIs use

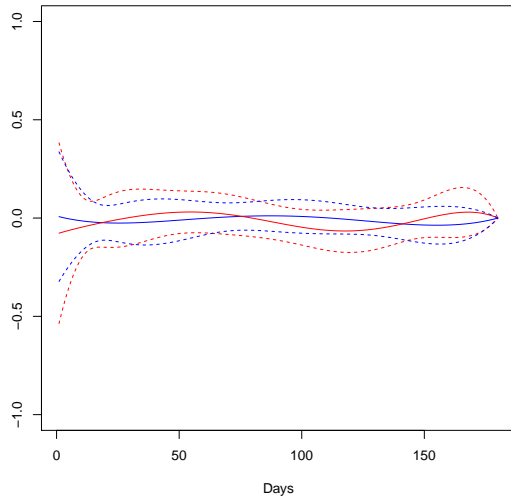
Exposure Windows	-2loglikelihood	Blood Pressure	Fixed Effects MEAN(SD)			$\hat{\sigma}_b$	$\hat{\sigma}_\varepsilon$
			Intercept	Time	Use of Hypertensive		
60	28342.00	Diastolic	72.25(5.30)	-0.003(0.0005)	8.30(4.63)	10.44	12.15
		Systolic	117.92(10.11)	-0.002(0.0009)	21.14(8.81)	20.48	20.03
120	29649.00	Diastolic	73.50(5.21)	-0.003(0.0005)	7.540(4.57)	10.39	12.14
		Systolic	118.87(9.90)	-0.002(0.0008)	21.70(8.70)	20.35	19.95
180	30456.00	Diastolic	72.24(5.26)	-0.002(0.0004)	8.04(7.54)	9.88	12.18
		Systolic	116.98(10.44)	-0.001(0.0008)	20.11(9.11)	21.05	20.04
360	32563.00	Diastolic	74.96(5.30)	-0.003(0.0005)	7.68(4.39)	10.56	12.03
		Systolic	116.38(9.81)	-0.002(0.0008)	22.15(8.40)	20.80	19.90



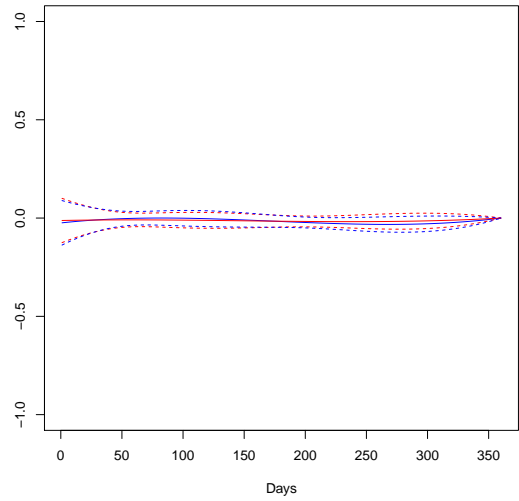
(a) Time window: 60 Days



(b) Time window: 120 Days



(c) Time window: 180 Days



(d) Time window: 360 Days

Figure 2.7: Estimated coefficient functions with corresponding confidence bands under various exposure time windows

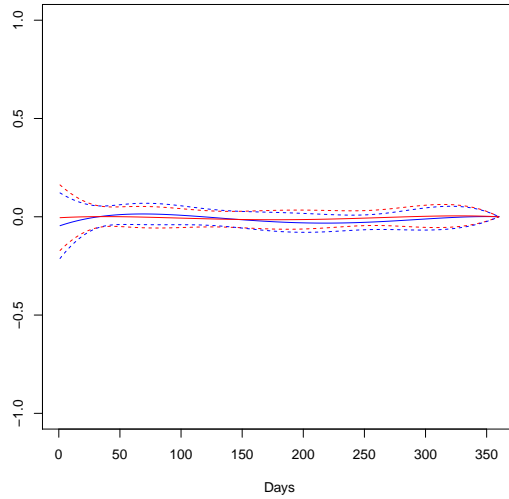


Figure 2.8: Estimated coefficient functions with corresponding confidence bands under exposure time window of 360 Days with 8 interior knots

## 2.6 Conclusion

In this Chapter, we developed functional regression models for multivariate outcomes with time-varying exposures using penalized splines. Our simulation studies indicate adequate performances for parameter estimation and inferences. Our proposed model extends previous methods in several ways. First, this model includes multivariate longitudinally measured outcomes and estimate the effects of exposure on all the outcomes simultaneously. Second, our model framework includes a constraint at the boundary of the exposure interval and hence reduces the estimation variability. By expressing the functional model into a mixed model representation, we demonstrate that parameter estimation and inference procedures can be implemented using standard statistical software packages.

We demonstrated the application of the proposed method in a group of SSRIs users in estimating the association between SSRIs exposure and longitudinal blood pressure measures. The proposed model framework can be readily applied to other medication groups

to determine whether medication use is associated with longitudinally measured health outcomes.

One limitation of this study is that we considered independent identical within-subject random errors for each type of outcome. However, a wider range of correlation structures could be embedded into within-subject random errors in the current modeling framework. For example,  $cor(\varepsilon_{ij}^{(k)}, \varepsilon_{ij'}^{(k)}) = q(|u_{ij} - u_{ij'}|)$ ,  $k = 1, \dots, K$ , where  $q$  is a correlation function taking values between -1 and 1. Such an assumption allows varying correlation over time but may introduce more computational complexity. Another direction for future work is to extend the coefficient function regression model to generalized mixed model settings, where the correlated outcomes can be from more general exponential distributions rather than normal distributions.

In summary, we proposed a functional regression model framework for multivariate outcomes with time-varying exposures using penalized spline. The proposed methods performed adequately in simulation studies. With the increasing availability of electronic medical records of medication dispensing data, our proposed methods can be readily applied to other medication dispensing data to determine whether medication use have a long-term association with longitudinally measured health outcomes.

## Chapter 3

### Flexible models for comparing cumulative effects of time-dependent exposures

#### 3.1 Introduction

One of the major focuses of pharmacoepidemiologic studies is to determine whether exposure to certain medications is associated with adverse health outcomes. The increasing use of electronic medical records (EMR) in health care systems has made patients' medication dispensing data available, including detailed information on doses and duration. In publications in the medical journals, many analyses were conducted using methods that summarize medication exposure by using current dose (Smitten et al., 2008), average daily dose (Wolfe et al., 2006) , or as simple sum of past exposures (Stranges et al., 2006). These methods all assume, explicitly or implicitly, that the impact of past exposure on event risk does not depend on the timing of the exposure. In addition, these methods do not combine the effects from both intensity and duration of time-varying exposures.

Breslow et al. (1983) and Thomas (1988) proposed the use of a weighted cumulative exposure (WCE) to summarize time varying exposures. Abrahamowicz et al. (2006) used the parametric WCE framework within the Cox's proportional hazards (PH) model to study the association between a time-varying exposure and event risk. They relied on a clinically viable weight function with more weight for recent exposures and gradually decreasing to zero weight thereafter, and a scale parameter for controlling the rate of change in weights over time. Interesting applications with weight function modeled by simple parametric functions can also be found in works by VACEK (1997) , Langholz et al. (1999), and Richardson (2009). However, without sufficient prior knowledge of the medication's pharmacokinetic properties, imposing a specific parametric form of the weight function may



result in invalid conclusion if the parametric form is mis-specified (VACEK, 1997). Instead of restricting the weight function to a parametric form, other approaches were proposed to estimate the functional form of the weight function using flexible nonparametric methods. Hauptmann et al. (2000) introduced a weight function model using cubic B-spline within generalized linear models to examine the time-dependent effects of smoking histories on lung cancer. A constrained maximization algorithm was then used to jointly estimate regression coefficient and spline coefficients. Sylvestre and Abrahamowicz (2009) extended the parametric WCE Cox model (Abrahamowicz et al., 2006) to nonparametric models using a regression spline-based method to assess the association between exposure to psychotropic drugs and fall-related injuries. The authors proposed re-parameterization of the WCE so that it can be implemented by survival analysis with time-dependent covariates in standard statistical packages.

Recently, many extensions have been made in defining WCE. For example, Berhane et al. (2008) proposed a tensor product spline model to jointly model age, latency and exposure response effects for protracted time-dependent occupational exposure histories in the Colorado plateau uranium miners cohort. Gasparrini (2013) noted out that equivalent approaches were previously established in time series analysis. Analogous to the weight function, the definition of a distributed lag function was introduced in distributed lag models (DLMs) and distributed lag nonlinear models (DLNMs) in time series analysis.

An existing problem, not addressed thus far is that in clinical practice patients receive different types of medications intended for the same medical condition and it is of interest to compare the effects of these medication exposures on relevant health outcomes. However, little work has been done in comparing the effect of multiple medication exposures over an extended period on time to event outcomes. In this Chapter, we extend the regression

spline-based method in Sylvestre and Abrahamowicz (2009) to model WCE in multiple exposures settings where weight functions for each exposure type can be considered.

In Section 3.2, we describe data from a primary care depression cohort as a motivating example for our proposed model. In Section 3.3, we present a flexible WCE model for multiple exposures on a survival outcome and describe procedures for parameter estimation, hypothesis testing and pointwise confidence bands for the weight functions. We present the results of a simulation study in section 3.4 to evaluate the performance of our proposed model. In Section 3.5, we apply the new method to the example data set and compare the risk of coronary artery disease (CAD) between two types of antidepressants. We include a discussion of our method in Section 3.6.

### **3.2 A primary care depression screening cohort**

Our interest is to compare the risk of CAD between two common types of antidepressants, tricyclic antidepressants (TCAs) and selective serotonin reuptake inhibitors (SSRIs) (Zellweger et al., 2004). There have been reports of increased risk of CAD events in TCA users (Pacher et al., 1999), but no increased CAD risk for SSRI users in postmenopausal women (Smoller et al., 2009). However, none of the previous studies considered dose or duration of antidepressants exposure in the analysis. As it is widely known that patients on depression treatments receive different dosages and may take these medications intermittently, it is therefore of interest to compare whether cumulative exposures to SSRIs and TCAs were associated with different CAD risks.

In our data set, there were 203 patients taking TCAs and 162 patients taking SSRIs during the follow-up time. From January 1991 to June 1993, those patients attending primary care clinics at the Wishard Health Services were enrolled into a depression screening study (Callahan et al., 1994). Electronic medical records on these patients from enrollment

to December 31, 2010 were extracted from the Regenstrief Medical Record System (RMRS) (McDonald et al., 1999). The RMRS is one of the first electronic medical record systems in the country and has been actively used for research purposes. Originally implemented in 1973 and used continuously since then, the RMRS serves as the day-to-day electronic medical record system at Wishard Hospital and its community clinics — the third largest safety-net health care system in the U.S. The RMRS routinely captures laboratory results, narrative reports, orders, medications, radiology reports, registration information, nursing assessments, vital signs, EKGs and other clinical data. The medical records also include comprehensive medication dispensing information capturing medication name, daily dose, beginning and ending dates for each dispensing record.

In this elderly population, patients were also at risk of dying. Patients who died of myocardial infarction (MI) without previous CAD should also be considered as CAD events. In order to combine such information, death dates were extracted from Medicare/Medicaid data set and causes of death were obtained from death certificates provided by the Indiana State Department of Health. Therefore, our primary outcome, a hard CAD event, was defined as the occurrence of any of the following events in the medical record or Medicare/Medicaid data during the follow-up period: (a) fatal MI; (b) laboratory evidence of acute MI (Creatine kinase-myocardial band isoenzyme value  $> 3.0$  ng/ml or troponin value  $> 0.3$  ug/L); and (c) diagnosis of CAD (ICD-9 codes 410-414,429.2).

To derive comparable dosages among different medications, we first standardized medication doses using daily dose divided by the recommended minimum dose for each medication (Damush et al., 2008). Medication dispensing data on antidepressants displays varied patterns for both dose levels and for the duration of medication use. In Figure 4.1, medication dosage and duration as well as event time from four randomly selected patients in the analysis cohort were displayed. Such patterns of medication exposure demonstrate the

need for statistical methods that can account for the varying exposure pattern and identify potential differences in the effects of exposure patterns over time on disease risk.

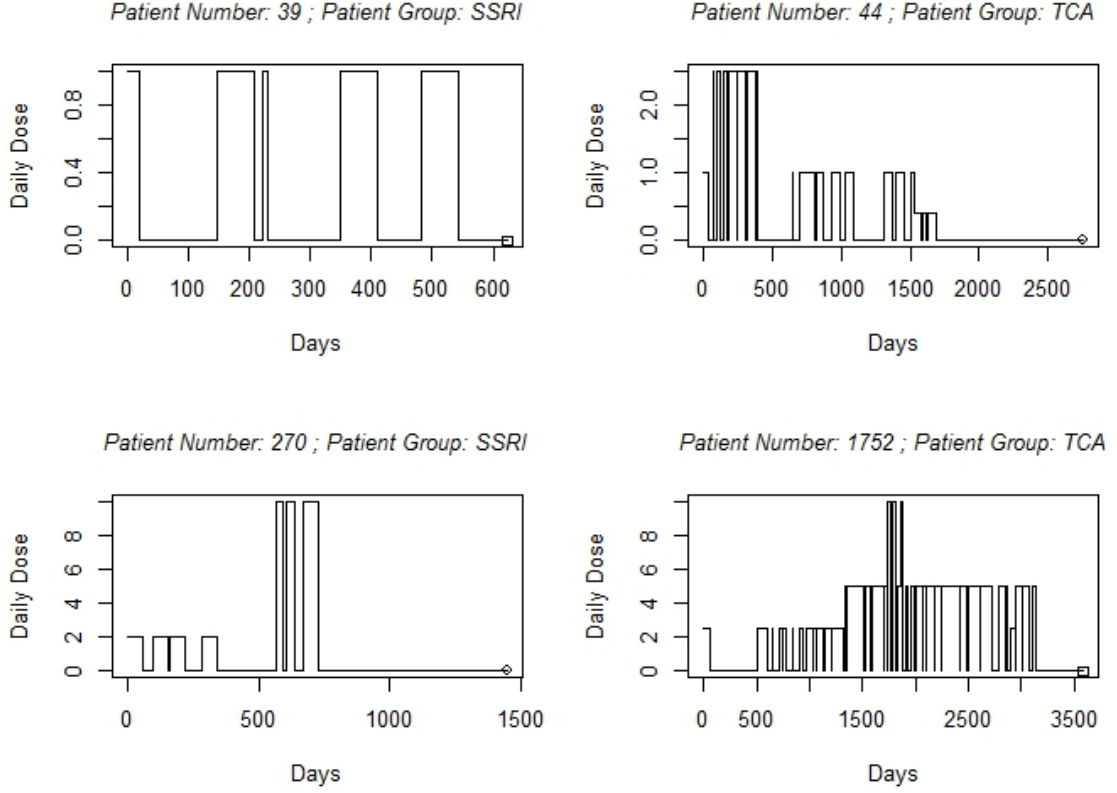


Figure 3.1: Medication exposure of four randomly selected patients with event time (denoted by square) or censoring time (denoted by diamond)

### 3.3 Method

#### 3.3.1 Cox's PH model with time-dependent WCE

Let  $X(t)$  be the exposure level measured at time  $t$ . Let  $M(u)$  be a summary measure of exposure up to time  $u$ :  $M(u) = \int_0^u f(X(t))dt$ , where  $f$  is a function that summarizes past exposures. The form of  $f$  reflects how the intensity and duration of past exposure are related to the outcome of interest.

The challenge is how to choose a functional form for  $f(\cdot)$  so that the summary exposure  $M(u)$  can best describe the cumulative effect of past exposures. A simple choice for the function  $f(\cdot)$  included the un-weighted cumulative dose, i.e.  $M(u) = \int_0^u X(t)dt$ , which assumes that past exposures have the same effect on the outcome no matter when the exposure happened.

Alternatively, a more flexible approach of summarizing past exposures can be achieved by assuming

$$M(u) = \int_0^u w(u-t)X(t)dt$$

where  $w(u-t)$  is an unknown weight function to be estimated from the data. Notice that the time axis in the weight function is defined as the length of time away from the endpoint, so that weight function defined on  $[0, a]$  would associate with history exposures within  $a$  units of time from an endpoint  $u$ . Without loss of generality, we assume that past exposure outside of the  $a$  units has negligible effect on the survival outcome by restricting the weight function to 0 for exposures beyond  $a$  unit length of time from the endpoint.

The weight function can be modeled using a polynomial function which will impose a strong prior assumption on the functional form. An alternative is to use splines with interior knots and make all pieces join smoothly to increase flexibility. The flexibility of the spline function is determined by the number of interior knots  $m$  and spline order  $q$ . Among different forms of spline bases, the B-spline (De Boor, 2001) is commonly used for its computational stability. In this paper, we use cubic B-splines ( $q = 4$ ) which provide sufficiently smooth estimates with continuous first two derivatives.

Once the number of interior knots of the cubic B-spline bases  $m$  is determined, the weight functions can be expressed as a linear combination of the B-spline bases:  $w(u-t) = \sum_{j=1}^{m+4} \theta_j B_j(u-t)$  (De Boor, 2001; Giorgi et al., 2003). Using such B-spline bases for the

weight function, the cumulative exposure to a medication can be written as

$$M(u) = \int_0^u \sum_{j=1}^{m+4} \theta_j B_j(u-t) X(t) dt$$

In real applications, the exposure  $X(t)$  is often measured at discrete time points. Therefore the cumulative exposure  $M(u)$  up to time  $u$  can be expressed as

$$M(u) = \sum_{t=0}^u \sum_{j=1}^{m+4} \theta_j B_j(u-t) X(t)$$

Suppose there are  $K$  medications of interest. Let  $I_k = I\{\text{med} = k\}$  be the indicator function for patients taking the  $k^{\text{th}}$  medication,  $k = 1, \dots, K$ , i.e.  $I_k = 1$  if a patient is on the  $k^{\text{th}}$  medication, and  $I_k = 0$  otherwise. The cumulative exposure which allows different forms of the weight functions for different medications can be written as:

$$M(u) = \sum_{t=0}^u \sum_{j=1}^{m+4} (\theta_{1j} + \sum_{k=2}^K \theta_{kj} I_k) B_j(u-t) X(t)$$

where  $\theta_{kj}, j = 1, \dots, m+4$  represent differences in weight functions between the  $k^{\text{th}}$  medication and the  $1^{\text{st}}$  medication group. Therefore,  $w_1(u-t) = \sum_{j=1}^{m+4} \theta_{1j} B_j(u-t)$  represents the weight function of the  $1^{\text{st}}$  medication group and  $w_k(u-t) = \sum_{j=1}^{m+4} (\theta_{1j} + \theta_{kj}) B_j(u-t)$  represents the weight function of the  $k^{\text{th}}$  ( $k > 1$ ) medication group.

If the outcome variable is time to certain event, a Cox's PH model can be fitted to assess the effect of the medication exposure on the event risk. The model involving medication exposure history as well as the types of medication can be written as:

$$h(u|\mathbf{X}(u), Z(u)) = h_0(u) \exp \left[ \beta_1 M(u) + \sum_{k=1}^K \beta_{2k} I_k + \sum_{s=1}^S \eta_s Z_s(u) \right] \quad (3.1)$$

where  $\beta_1$  represents the effect of the weighted cumulative exposure on the hazard in people who took the first medication.  $\beta_{2k}$ ,  $k = 1, \dots, K$  represents the medication difference compared to the reference group adjusting for other covariates in the model. For parameter identifiability, we set  $\beta_{21} = 0$ , which in essence sets the first medication group as the reference group.  $Z_s(u)$ ,  $s = 1, \dots, S$  are other covariates. Notice that, if  $K = 1$  then our model is consistent with the WCE model proposed by Sylvestre and Abrahamowicz (2009).

### 3.3.2 Estimation procedures

The estimation of the survival model (3.1) can be implemented by introducing time-dependent covariates  $D_j(u) = \sum_{t=0}^u B_j(u-t)X(t)$ ,  $j = 1, \dots, m+4$ , where  $D_j(u)$  can be thought of as the exposure weighted by a set of predetermined basis functions  $B_j(u-t)$ . Therefore, the proposed model (3.1) using the variables  $D_j$ , medication main effect and interactions can be rewritten as

$$\begin{aligned} h(u|\mathbf{X}(u), Z(u)) = h_0(u) \exp & \left[ \sum_{j=1}^{m+4} \beta_1 \theta_{1j} D_j(u) + \sum_{k=2}^K \beta_{2k} I_k \right. \\ & \left. + \sum_{k=2}^K \sum_{j=1}^{m+4} \beta_1 \theta_{kj} D_j(u) * I_k + \sum_{s=1}^S \eta_s Z_s(u) \right] \end{aligned} \quad (3.2)$$

Notice that  $\beta$ 's and  $\theta$ 's cannot be estimated separately from the above model. Instead, their products:

$$\gamma_{kj} = \beta_1 \theta_{kj}, k = 1, \dots, K$$

are estimable from the model. The new parameters combine both shape and scale information in the weight functions into a single parameter.

Correspondingly, model (3.2) becomes

$$\begin{aligned}
h(u|\mathbf{X}(u), Z(u)) = & h_0(u) \exp\left[\sum_{j=1}^{m+4} \gamma_{1j} D_j(u) + \sum_{k=2}^K \beta_{2k} I_k \right. \\
& \left. + \sum_{k=2}^K \sum_{j=1}^{m+4} \gamma_{kj} D_j(u) * I_k + \sum_{s=1}^S \eta_s Z_s(u)\right]
\end{aligned} \tag{3.3}$$

With a scale difference of  $\beta_1$ , the weight function for the 1<sup>st</sup> medication group can now be represented as  $w_1(u - t) = \sum_{j=1}^{m+4} \gamma_{1j} B_j(u - t)$  and  $w_k(u - t) = \sum_{j=1}^{m+4} (\gamma_{1j} + \gamma_{kj}) B_j(u - t)$  for the  $k^{th}$  ( $k > 1$ ) medication group.

With the new time-varying covariates,  $D_j(u)$ , model (3.3) can be fitted using procedures for Cox's PH model with time dependent covariates in any standard statistical software. As we mentioned before, when the order of splines is fixed, the flexibility of the estimated spline function is determined by the number of interior knots. Akaike information criterion(AIC) (Akaike, 1974) can be used to select the model with best predictive performance from models with different numbers of interior knots. The AIC adapted to a Cox PH model is defined as:  $AIC = -2\ln(PL) + 2p$ , where  $PL$  is the partial likelihood and  $p$  is the number of estimable parameters (Abrahamowicz et al., 2012). The model with the smallest AIC is selected as the best fitting model to the data at hand. Comparisons of WCE models using the minimum AIC criteria were shown to identify the true model in the vast majority of simulation studies conducted by Abrahamowicz et al. (2012).

### 3.3.3 Hypothesis testing

In pharmacoepidemiologic studies, multiple medications are often available for the same medical condition and thus the comparison of adverse event risk among different medication users is of great interest. Several hypotheses may be proposed for model (3.3). The first hypothesis is whether the  $K$  medications share the same weight function, *i.e.*, whether



past exposures influence current risk in the same way for all medications. Based on model (3.3), linear combinations of B-spline bases with coefficients  $\gamma_{kj}$ ,  $j = 1, \dots, m + 4$  represent differences in weight functions between the  $k^{th}$  medication and the reference group, where  $k = 2, \dots, K$ .

Therefore, the hypothesis of no difference in weight function can be stated as:

$$H_0: \gamma_{kj} = 0, \quad k = 2, \dots, K, \quad j = 1, \dots, m + 4$$

$$H_1: \textit{otherwise}$$

where  $m$  is the number of interior knots within the time interval  $[0, a]$ .

Under the assumption that the spline basis is chosen as a prior, it had been shown that in Cox PH models the partial likelihood score statistics are asymptotically normal, hence Likelihood Ratio Test (LRT) using partial likelihood functions can be used to test this hypothesis since the model under the null hypothesis of equal weight function is nested within the full Model (3.3) (Andersen and Gill, 1982). Asymptotic normality of model parameters was also previously established (Andersen and Gill, 1982). However, when model selection is used, the LRT statistics are conditional on the selected model, where the degree of freedom of the Chi-square statistic is no longer fixed and depends on the number of selected interior knots. Standard LRT ignoring model selection may lead to inflated type I error (Abrahamowicz et al., 1996; Hurvich and Tsai, 1990; Mahmud et al., 2006). We will assess the impact of the violation to the fixed-basis assumption in simulation studies in Section 3.4.

Another hypothesis of interest is whether the medications have the same main effect, i.e.  $\beta_{2k} = 0$ ,  $k = 2, \dots, K$ . Wald tests using the asymptotic normality of the partial maximum likelihood estimates can be used to test these hypotheses.

### 3.3.4 Pointwise confidence band for the estimated weight function

In order to evaluate the variation of the estimated weighted functions, nonparametric bootstrap re-sampling can be used to construct 95% confidence band of estimated weight function (Efron, 1979). Subjects can be re-sampled with replacement with their medication exposure history and endpoint from the original data set to create re-sampled data sets of the same size. For each of the bootstrap samples the proposed WCE models can be fitted with the same number of interior knots as in the original data set. The empirical distribution of the point estimates of weight function is constructed and 95% pointwise confidence bands can be computed using the bootstrap samples. The confidence band using bootstrap re-sampling accounts for the variation of regression coefficients as well as the additional variation introduced by the selection of knots.

## 3.4 Simulation study

To evaluate the performance of our proposed model, we simulated data sets using a prospective cohort design to compare the effects of two different medications on the risk of an adverse event. We considered two different sample sizes (200 and 500 patients) with equal probability of taking one of two medications.

### 3.4.1 Generating medication exposures and survival data

Medication exposures were generated in a one-year follow-up window with days being the unit of time. The duration of treatment period or interruption periods, in 7 days intervals, was generated from a lognormal distribution with mean 0.5 and standard deviation 0.8 on the log scale, i.e.  $\log(duration) \sim N(0.5, 0.8^2)$ , and rounded up to integer values. Daily doses were assumed to be constant over each treatment period but can vary during the follow up period with equal probability of 0.5, 1, 1.5, 2, 2.5, or 3. We also generated a

time-fixed variable for age at baseline from normal distribution with mean 70 and variance 3.

We considered the combinations of six true weight functions for medications,  $A$  and  $B$ , with a pre-specified time window of 180 days (Figure 3.2), similar to the weight functions used by Sylvestre and Abrahamowicz (2009) in the single exposure setting . The six scenarios for weight functions include maximum medication effect closer to event time with decreasing effect for more distant exposures (Scenarios 1 and 2), peak effect followed by decreasing effect (Scenario 3), increasing effect followed by decreasing trend in an inverted U function (Scenario 4), constant effect (Scenario 5) and increasing effect for distant exposures (Scenario 6). The combinations of these six scenarios for each of the two medications create 21 unique weight function patterns for the two medications. For convenience, we denote the combination of scenario  $i$  for Medication A and scenario  $j$  for Medication B as  $A_iB_j$ , where  $i, j = 1, \dots, 6, j \geq i$ .

Time to event was generated using the permutation algorithm designed and validated for simulating survival times conditional on time-dependent covariates (Sylvestre and Abrahamowicz, 2008).

### 3.4.2 Simulation results

We considered models with the number of interior knots ranging from one to three, which implies five to seven bases using cubic B-spline bases. For each of the 1000 samples simulated under a given scenario, we used the AIC criteria to select the number of knots for the model.

#### Assessing parameter estimates

In our simulation setting with two medications, we drop the subscript  $k$  of main effect  $\beta_{2k}$  in our models. The estimable parameters from model (3.3) include  $\beta_2$ ,  $\eta_1$ , and  $\gamma_{kj}$ ,  $k = 1, 2; j = 1, \dots, m + 4$ . In order to estimate the coefficients of WCE, we use normalized

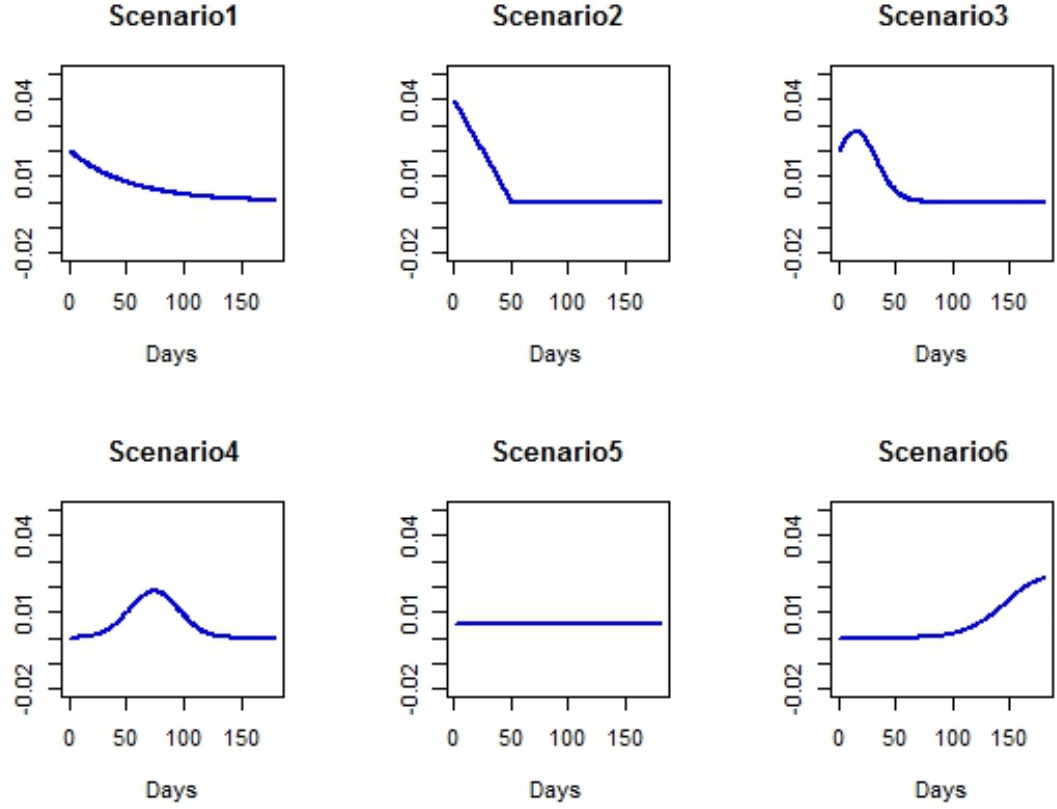


Figure 3.2: Six scenarios of true weight function for each medication. The time axis is the reversed time since exposure. The origin  $Days = 0$  corresponds to current time, and increasing  $Days$  correspond to more distant past. All the weight functions are normalized.

weight functions:

$$\sum_{\tau=0}^a w(\tau) = 1$$

we can estimate the effect of WCE as:

$$\hat{\beta}_1^1 = \sum_{\tau=0}^a \sum_{j=1}^{m+4} \hat{\gamma}_{1j} B_j(\tau)$$

and

$$\hat{\beta}_1^2 = \sum_{\tau=0}^a \sum_{j=1}^{m+4} (\hat{\gamma}_{1j} + \hat{\gamma}_{2j}) B_j(\tau)$$

Parameter estimates are presented in Table 3.1. For a given scenario, average parameter estimate, estimated standard deviation and empirical standard deviation from 1000 simulations were presented for each coefficient. Since  $\exp(\beta_2)$  represents the hazard ratio between medication A and medication B given the same cumulative exposure, we also reported 95% confidence coverage rate for  $\beta_2$ .

Table 3.1 shows that the estimated coefficients are unbiased across all scenarios in both sample sizes. As expected, bias and standard deviation estimates, both model-based and empirical based, decrease when sample size increases. The estimated standard deviation closely tracks the empirical standard deviation with both sample sizes.

Figure 3.3 displays a random sample of 100 estimated weight functions with the corresponding true weight functions in thick white. Each estimated weight function was calculated based on the selected model using AIC criteria for a given simulated data set. In most scenarios, the estimated weight functions were able to capture the shape of the true weight functions and can be distinguished between the two medication groups. As expected, the estimate weight functions show considerable over-fitting bias in the tails, which is a known feature of B-splines (Hastie and Tibshirani, 1986).

### **Hypothesis testing on weight functions**

Table 3.2 presents the proportion of the 1000 simulations when the LRT rejects the null hypothesis of no difference in weight functions ( $p < 0.05$ ) for the 21 unique weight function combinations. Therefore, entries on the diagonal represent estimated type I error rates and entries in the off-diagonal setting are empirical power of the LRT. Two different rejection rates were calculated under each scenario. In the first set of calculation, we fixed the number of interior knots to three, while in the second method, the degree of freedom of the LRT statistics depends on the number of interior knots selected using AIC.

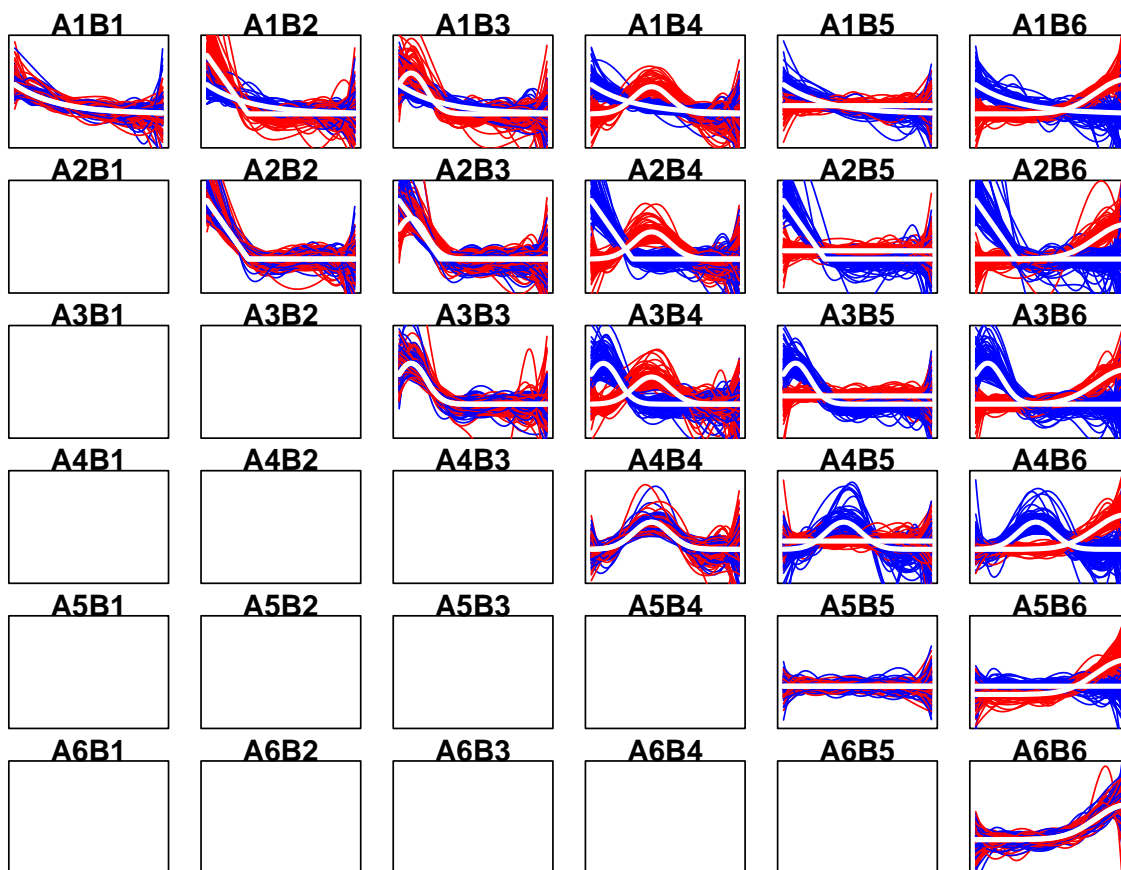


Figure 3.3: A random sample of 100 normalized estimated weight functions (Sample size=500). Thick white lines indicate the true weight functions . Blue lines indicate weight functions for Medication A and red lines indicate weight functions for Medication B.

Simulation results show that empirical type I error rates for the fixed method are closer to the nominal level of  $\alpha = 0.05$  and empirical type I error rates using the flexible method are slightly inflated. Furthermore, in almost all the situations, estimated power is high for both methods except for scenario *A2B3*, where the two weight functions were very similar (Figure 3.2).

### Assessing pointwise confidence band

Without loss of generality, we assess the precision of the estimated weight function using confidence band under the null hypothesis of no difference between the two medications.

Figure 3.4 shows four randomly selected bootstrap confidence bands when both medications

have inverted U weight functions (Scenario *A4B4*). Solid lines indicate true weight functions and dotted ones indicate estimated weight functions from the original simulated data set. Each panel presents the estimated upper and lower 95% confidence band (dashed line). Coverage proportion (CP), which is the percentage of equally spaced sampling points where the confidence bands cover the true weight function, was also calculated (Wahba, 1983). Average coverage proportion (ACP) over 100 simulated data sets was 94.6%, close to the nominal level of 95% .

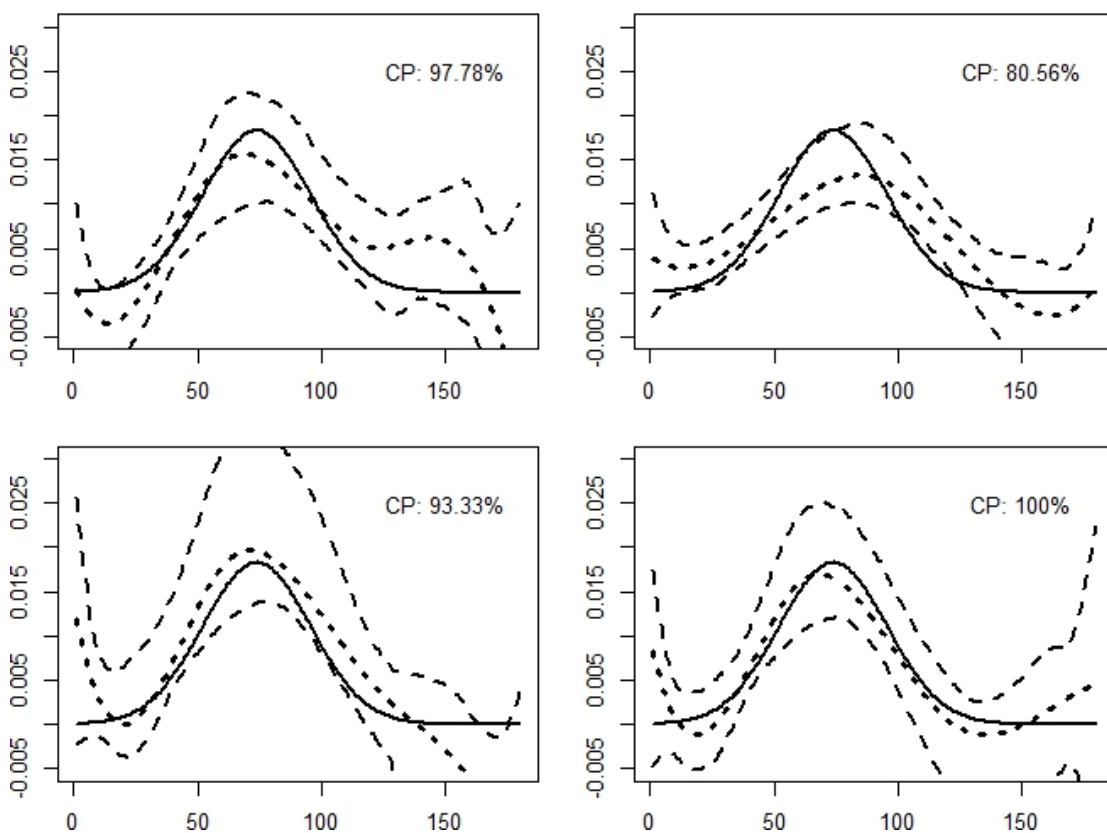


Figure 3.4: Four randomly selected bootstrap confidence bands of inverted U weight function (Scenario *A4B4*, Sample size=500): True weight function used to simulated data is denoted by solid line; Dotted line is the estimated weight function from original simulated data set; Dashed lines are upper or lower bound of 95% bootstrap confidence band

Table 3.1: Estimated coefficients based on 1000 simulations with sample size 200 and 500

Sce.	$\beta_1^1$ (True 2.0)			$\beta_1^2$ (True 2.0)			$\eta_1$ (True 0.5)			$\beta_2$ (True 1.0)			
	M.	M.SD	E.SD	M.	M.SD	E.SD	M.	M.SD	E.SD	M.	M.SD	E.SD	Cov.R
Sample size 200													
A1B1	2.02	0.69	0.74	2.03	0.76	0.80	0.52	0.05	0.05	0.99	0.48	0.51	0.94
A1B2	1.96	0.69	0.72	1.98	0.84	0.88	0.51	0.05	0.05	1.00	0.52	0.56	0.94
A1B3	2.02	0.69	0.72	2.03	0.83	0.88	0.52	0.05	0.05	1.02	0.51	0.54	0.94
A1B4	1.98	0.70	0.75	1.97	0.77	0.82	0.52	0.05	0.05	1.02	0.47	0.51	0.93
A1B5	2.01	0.71	0.74	2.02	0.71	0.74	0.52	0.05	0.05	1.02	0.47	0.49	0.95
A1B6	2.04	0.75	0.77	2.08	0.67	0.69	0.52	0.05	0.05	1.04	0.47	0.48	0.95
A2B2	2.04	0.74	0.79	2.04	0.81	0.86	0.52	0.05	0.05	1.04	0.53	0.55	0.94
A2B3	2.06	0.74	0.77	2.04	0.80	0.84	0.52	0.05	0.05	1.02	0.52	0.55	0.94
A2B4	2.03	0.77	0.80	2.01	0.77	0.81	0.52	0.05	0.05	1.05	0.51	0.54	0.94
A2B5	2.04	0.77	0.83	2.01	0.70	0.75	0.52	0.05	0.05	1.06	0.51	0.50	0.96
A2B6	2.02	0.81	0.84	2.04	0.67	0.70	0.52	0.05	0.05	1.06	0.51	0.53	0.95
A3B3	2.01	0.73	0.75	1.99	0.80	0.84	0.52	0.05	0.05	1.04	0.52	0.54	0.93
A3B4	1.99	0.75	0.76	2.01	0.77	0.79	0.52	0.05	0.05	1.01	0.50	0.51	0.94
A3B5	2.05	0.77	0.82	2.03	0.71	0.77	0.52	0.05	0.05	1.07	0.50	0.51	0.94
A3B6	1.99	0.80	0.85	2.00	0.67	0.71	0.52	0.05	0.05	1.05	0.50	0.52	0.95
A4B4	2.06	0.70	0.79	2.05	0.78	0.84	0.52	0.05	0.05	1.01	0.46	0.48	0.94
A4B5	2.04	0.70	0.76	2.03	0.72	0.79	0.51	0.05	0.05	1.05	0.47	0.50	0.94
A4B6	2.04	0.74	0.78	2.04	0.66	0.70	0.52	0.05	0.05	1.06	0.46	0.48	0.94
A5B5	2.02	0.65	0.68	2.04	0.73	0.77	0.52	0.05	0.05	1.02	0.47	0.48	0.95
A5B6	1.97	0.67	0.73	2.02	0.66	0.70	0.51	0.05	0.05	1.02	0.46	0.49	0.94
A6B6	2.05	0.63	0.63	2.07	0.68	0.69	0.52	0.05	0.05	1.03	0.46	0.47	0.95
Sample size 500													
A1B1	2.00	0.41	0.42	2.00	0.44	0.45	0.51	0.03	0.03	1.01	0.31	0.32	0.94
A1B2	2.01	0.42	0.41	2.01	0.48	0.50	0.51	0.03	0.03	1.01	0.34	0.35	0.96
A1B3	2.00	0.42	0.43	2.01	0.47	0.48	0.51	0.03	0.03	1.01	0.33	0.34	0.94
A1B4	2.01	0.42	0.44	2.02	0.43	0.45	0.51	0.03	0.03	1.02	0.29	0.29	0.95
A1B5	2.02	0.42	0.43	2.03	0.40	0.41	0.51	0.03	0.03	1.02	0.29	0.29	0.95
A1B6	1.98	0.45	0.45	2.01	0.37	0.37	0.51	0.03	0.03	1.02	0.29	0.29	0.95
A2B2	2.02	0.44	0.43	2.01	0.46	0.48	0.51	0.03	0.03	1.04	0.34	0.35	0.94
A2B3	2.03	0.44	0.44	2.06	0.46	0.46	0.51	0.03	0.03	1.01	0.33	0.33	0.95
A2B4	2.03	0.46	0.46	2.02	0.43	0.43	0.51	0.03	0.03	1.04	0.31	0.31	0.95
A2B5	2.00	0.45	0.47	1.99	0.39	0.41	0.50	0.03	0.03	1.02	0.31	0.32	0.94
A2B6	2.00	0.49	0.51	2.01	0.37	0.39	0.51	0.03	0.03	1.03	0.30	0.30	0.96
A3B3	2.01	0.44	0.43	2.00	0.46	0.49	0.51	0.03	0.03	1.01	0.33	0.33	0.95
A3B4	2.00	0.45	0.47	1.99	0.43	0.43	0.51	0.03	0.03	1.02	0.30	0.32	0.94
A3B5	2.01	0.45	0.46	2.01	0.39	0.41	0.51	0.03	0.03	1.03	0.30	0.31	0.94
A3B6	2.01	0.48	0.48	2.01	0.37	0.37	0.51	0.03	0.03	1.03	0.30	0.30	0.96
A4B4	2.01	0.42	0.43	2.01	0.44	0.46	0.51	0.03	0.03	1.02	0.29	0.30	0.94
A4B5	2.00	0.43	0.43	2.01	0.41	0.43	0.51	0.03	0.03	1.01	0.29	0.29	0.96
A4B6	2.00	0.45	0.45	2.02	0.37	0.37	0.51	0.03	0.03	1.01	0.29	0.29	0.95
A5B5	2.00	0.39	0.39	2.01	0.41	0.43	0.51	0.03	0.03	1.01	0.29	0.31	0.93
A5B6	2.01	0.41	0.42	2.02	0.37	0.38	0.51	0.03	0.03	1.02	0.28	0.29	0.94
A6B6	2.04	0.37	0.39	2.04	0.38	0.39	0.51	0.03	0.03	1.01	0.29	0.29	0.95

M. is the mean of estimates across 1000 simulation results

M.SD is the mean of model based standard deviation across 1000 simulation results

E.SD is empirical standard deviation of 1000 estimated coefficients

Cov.R is the coverage rate of 95% confidence intervals across 1000 simulation results



Table 3.2: Hypothesis test of equal weight functions (sample size=500). Entries in the table are proportion of simulations where the null hypothesis is rejected based on 1000 simulations

		<i>Medication<sub>B</sub></i>					
<i>Medication<sub>A</sub></i>		S.1	S.2	S.3	S.4	S.5	S.6
$\beta_2 = 0$							
S.1	Fixed	0.065	0.968	0.859	0.998	0.845	1.000
	Flexible	0.082	0.983	0.884	0.999	0.896	1.000
S.2	Fixed		0.048	0.288	1.000	1.000	1.000
	Flexible		0.065	0.330	1.000	1.000	1.000
S.3	Fixed			0.050	1.000	1.000	1.000
	Flexible			0.050	1.000	1.000	1.000
S.4	Fixed				0.041	0.933	1.000
	Flexible				0.068	0.962	1.000
S.5	Fixed					0.067	0.973
	Flexible					0.057	0.988
S.6	Fixed						0.046
	Flexible						0.086
$\beta_2 = 1$							
S.1	Fixed	0.060	0.962	0.832	0.999	0.851	1.000
	Flexible	0.071	0.974	0.853	0.999	0.893	1.000
S.2	Fixed		0.046	0.301	1.000	1.000	1.000
	Flexible		0.073	0.354	1.000	1.000	1.000
S.3	Fixed			0.051	1.000	1.000	1.000
	Flexible			0.069	1.000	1.000	1.000
S.4	Fixed				0.065	0.953	1.000
	Flexible				0.058	0.966	1.000
S.5	Fixed					0.045	0.978
	Flexible					0.064	0.992
S.6	Fixed						0.063
	Flexible						0.066

### 3.5 Data application

In this section we revisit the example data we described in Section 3.2 and apply the proposed model to the data from the primary care depression cohort. Our interest is to compare the risk of CAD between patients on SSRIs to those on TCAs. Patients with CAD before antidepressant use were excluded from the analysis. Therefore, the analysis data set contained 365 subjects in total with 203 patients taking TCAs and 162 patients on SSRIs. Entry time in the survival analysis was the first time when a patient was dispensed an antidepressant (Suissa, 2007). Event time was calculated using CAD date if a patient had CAD or death date if a patient died of CAD. Censoring time was calculated using last clinical encounter date for those who were alive or death data if a patient died of other disease. A total of 123 patients (33.7%) developed hard CAD during the follow-up.

The median survival time is 1448 days in this sample. In the WCE model framework, we considered varying windows of medication effects from 60 days up to 1920 days with the length of window doubling each time. For each pre-specified time window, we estimated WCE models with three fixed interior knots uniformly spaced within the time window. In addition to medication variables, all models also adjusted for five covariates for patients' characteristics: age at first antidepressant use, years of education, sex, race and a binary indicator for smoking status.

Table 3.3 summarizes the corresponding AIC as well as p-values from hypothesis tests for each pre-specified time window. The model with 240 days of potential medication effect had the smallest AIC. Regardless, results from LRT and Wald tests for all the models indicate that there is no difference between the two weight functions and there is also no difference between the main effect of SSRIs and TCAs.

Figure 3.5 shows the estimated weight functions with 240-day window for both SSRIs and TCAs user as well as their 95% bootstrap pointwise confidence bands. In general,

Table 3.3: Model results using different medication exposure windows from 60 days up to 1920 days

Time window (Days)	AIC	Hypothesis test of no weight function difference	Hypothesis test of no medication main effect
60	1438.91	0.61	0.99
120	1437.97	0.52	0.91
240	1434.96	0.29	0.94
480	1436.90	0.44	0.99
960	1438.10	0.51	0.80
1920	1439.19	0.31	0.76

the estimated weight functions of SSRIs and TCAs are similar across the entire time window, which is consistent with the hypothesis test results. Both 95% pointwise confidence bands include zero across the entire exposure window. Overall, in this cohort, we found no difference in CAD risk between patients taking SSRIs and those taking TCAs.

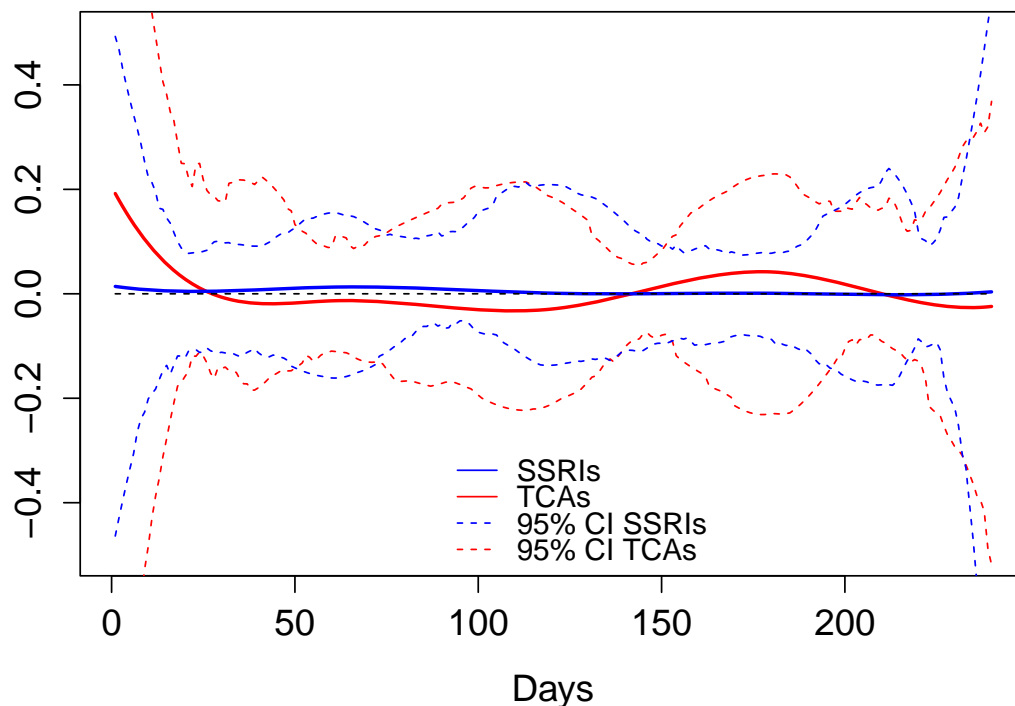


Figure 3.5: Estimated weight functions with 240-day window for antidepressant SSRIs and TCAs user as well as 95 percent bootstrap confidence band: solid lines indicate estimated weight functions and dashed lines indicate 95% confidence bands (SSRIs in blue and TCAs in red); Dashed horizontal line is the zero line.

### 3.6 Conclusion

In observational studies, especially in post-marketing surveillance studies monitoring medications for long-term adverse events, it is important to utilize patients' medication history when comparing different classes of medications for the same indication. In this Chapter, we extend the flexible WCE model proposed by Sylvestre and Abrahamowicz (2009) for single exposure to the comparison of multiple exposures. The use of flexible weight functions using spline bases not only offers flexibility in summarizing cumulative exposures, but also provides information on how the risk of adverse event is impacted by the timing of the exposures. Furthermore, parameter estimation for the proposed model can be

implemented using standard statistical packages for survival analysis with time-dependent covariates making the proposed models applicable to many data applications.

Our simulation studies demonstrate adequate performance of the proposed methods in parameter estimation as well as hypothesis testing using fixed number of knots. However, simulation results also revealed inflated Type I error rates when the test was conditional on AIC criteria selected number of knots. Further research is needed to determine a more appropriate distribution for the LRT statistics when using a model selection criteria to determine the number of knots.

Some extensions can be made to the proposed methods in future studies. The models considered here did not use any constraint on the weight function when using spline bases. As a result, the estimated weight functions can be negative as seen in the simulations and the analysis of the antidepressant data. In practice, such negative weight functions are difficult to interpret and it would be ideal to start the model framework with the non-negative requirement in the model set-up. Parameter estimation and hypothesis testing procedures need to be modified to accommodate such a constraint.

In aging studies involving patients with multiple chronic conditions, the majority of patients were found to be on multiple medications at any given time. Thus it is important that statistical models take into account all medications a patient has been exposed in order to detect potential risk to adverse events. Another significant development based on our methods is to model interactions among the different medications patients are taking.

In summary, we proposed a flexible model to compare cumulative exposures of multiple medications over time. Parameter estimation and hypothesis tests can be conducted using standard statistical software packages for survival analysis. The proposed methods performed adequately in simulation studies. With the increasing availability of electronic medical records of medication dispensing data, our proposed methods can be readily applied

to medication dispensing data to detect differences in adverse event risk among different classes of medications.

## Chapter 4

### A penalized Cox PH model with time-varying exposures and interaction term

#### 4.1 Introduction

In recent pharmacoepidemiological studies, the increasing use of electronic medical records (EMR) in health care systems has made available patients medication dispensing data with detailed information on medication names, dosage and duration. Such particular medication exposure poses challenges to existing statistical model because both the exposure status and its intensity vary over time, named as time-varying exposure in general. Moreover, patients often receive multiple types of medication, sequentially or concurrently, intended for the same medical condition. For example, depressed patients may switch from one type of antidepressants to another if the treatment fails to show satisfactory efficacy. Or patients with hypertension may take more than one class of antihypertensive medications at the same period. Therefore, assessing the effect of multiple exposures and their potential interaction effects brings additional complexity to the statistical model.

To fully utilize the information of such time-varying exposure, Breslow et al. (1983) and Thomas (1988) were among the first ones to propose the concept of weighted cumulative exposure (WCE) by assigning predetermined weights to time-varying exposures and summarizing it into scalars. Applications using WCE by pre-defined parametric functions can also be found in works by Abrahamowicz et al. (2006), VACEK (1997), Langholz et al. (1999), and Richardson (2009). Various approaches have also been proposed using nonparametric splines to estimate the weight function in generalized linear models (Berhane et al., 2008; Hauptmann et al., 2000) and Cox proportional hazard models (Gasparrini, 2013; Sylvestre and Abrahamowicz, 2009). However, to the best of our knowledge, few study discussed

multiple time-varying exposures, let alone the interactions between them, in WCE model setting.

In the meanwhile, functional data analysis is also an active field to study time-varying exposures. Functional regression models have been developed for functional covariates, where the functional coefficients can be equivalent to the weights for time-varying exposures in weighted cumulative exposure models. Many functional regression models include only one single functional covariate, such as Zhang et al. (2007), Schipper et al. (2008) and Bhadra et al. (2012). Goldsmith et al. (2012) and Ferraty and Vieu (2009) include two or more functional covariates through additive models. If the assumption of additivity of the effects of multiple functional covariates is challenged, it is natural to add covariate interaction effects. To the best of our knowledge, not much work has been done to study the interaction effects between functional covariates. Fuchs et al. (2015) generalized functional linear model and proposed a penalized scalar-on-functions regression with interaction term, which extended the model with only main effects in Wood (2011). The estimation of interaction effects between functional covariates has not received much attention in the survival setting.

Additionally, when modeling the effect of exposure to medication or environmental toxins, there are biological reasons that exposures occurring long time before an outcome might have negligible influence, so that the coefficient function should be bounded toward zero at a distant boundary. Schipper et al. (2008) proposed two constraints on the coefficient function making the coefficient function zero at zero dose and imposing monotonicity such that higher dose leads to more adverse outcomes.

In this chapter, we proposed a penalized Cox PH model with functional covariates and interaction term. In particular, the coefficient functions for functional covariates have been expanded using cubic B-spline basis for the main effects and tensor product splines for the



interaction effect. Because boundary issues may happen in estimating the coefficient functions when regression spline was used (Sylvestre and Abrahamowicz, 2009), ridge penalties were added to smooth coefficient functions for the main effect and coefficient surface for the interaction effect. Studies about penalized Cox PH model and its associated hypothesis tests can be found in Gray (1992, 1994), Therneau et al. (2003) and Perperoglou (2014). The coefficient functions and coefficient surface were estimated by maximizing penalized log partial likelihood. Hypothesis tests of no interaction effect or no effect of time-varying exposure were also discussed based on Wald’s method. The performance of the model is evaluated in simulations.

This chapter is organized as follows. Section 2 describes data from a primary care depression study as a motivating example. Section 3 introduces the penalized Cox PH model with interaction term, parameter estimation and hypothesis test. Section 4 presents results of simulation studies. In section 5, we apply the proposed model to the real data set and estimate the association between two types of time-varying antidepressant exposure and the risk of coronary artery disease.

## **4.2 A primary care depression screening cohort**

From January 1991 to June 1993, 3767 elderly patients attending primary care clinics at the Wishard Health Services were enrolled into a depression screening study (Callahan et al., 1994). Electronic medical records on these patients from enrollment to December 31, 2010 were extracted from the Regenstrief Medical Record System (RMRS) (McDonald et al., 1999). The RMRS is one of the first electronic medical record systems in the country and has been actively used for research purposes. The RMRS routinely captures laboratory results, narrative reports, orders, medications, radiology reports, registration information, nursing assessments, vital signs, EKGs and other clinical data. The medical records include

comprehensive medication dispensing information capturing medication name, daily dose, beginning and ending dates for each dispensing record.

Antidepressant is one of the most commonly prescribed medication groups in the United States (Lindsley, 2012). An older class of antidepressants, tricyclic antidepressants (TCAs), have been shown to have detrimental effect on cardiovascular function by inhibiting cardiovascular  $\text{Na}(+)$ ,  $\text{Ca}(2+)$  and  $\text{K}(+)$  channels often leading to life-threatening arrhythmia (Glassman, 1984; Jefferson, 1975). A newer class of antidepressants, selective serotonin reuptake inhibitors (SSRIs), became the preferred treatment for depression due to its comparable efficacy with TCAs and its superior tolerability. SSRIs were hypothesized to show a different cardiovascular effect from TCAs due to their pharmacologic profile (Bergstrom et al., 1988). However, there have been reports of first-degree atrioventricular block, prolonged QTc interval, and orthostatic hypotension in SSRI-treated patients suggesting that SSRIs may also have important cardiac and vascular effects (de la Torre et al., 2001; Pacher and Kecskemeti, 2004). Given that many of these studies were conducted in the laboratory setting with brief antidepressant treatment, EMR data with detailed medication dispensing information offer a unique opportunity to examine the effect of long-term use on cardiac functions in an elderly patient population. Moreover, none of the previous studies considered the comprehensive information of time-varying antidepressants exposure in the analysis. As it is widely known that patients on depression treatments receive different dosages and may take these medications sequentially, it is therefore of interest to examine whether the time-varying exposures to TCAs or SSRIs and their interaction are associated to risk of CAD.

In the data set, there were 930 depression patients with at least one dispensing record for antidepressants. We included patients who were taking TCA or SSRI throughout the follow-up period. The time when they initiated the medication was considered as time

zero in our survival analysis. To derive comparable dosages among different medications, we first standardized medication doses using daily dose divided by the recommended minimum dose for each medication (Damush et al., 2008). Dispensing data of antidepressants display various patterns for both dose intensity and duration of treatment. In Figure 4.1, medication dosage and duration from four typical patients in the analysis cohort were displayed: patient who took TCA exclusively throughout the follow-up period; patient who take SSRI exclusively; patient who switched from TCA to SSRI; patient who took TCA and SSRI sequentially. Such patterns of medication exposure demonstrate the need for statistical methods that can account for the varying exposure pattern and identify potential differences in the effects of exposure patterns over time on disease risk.

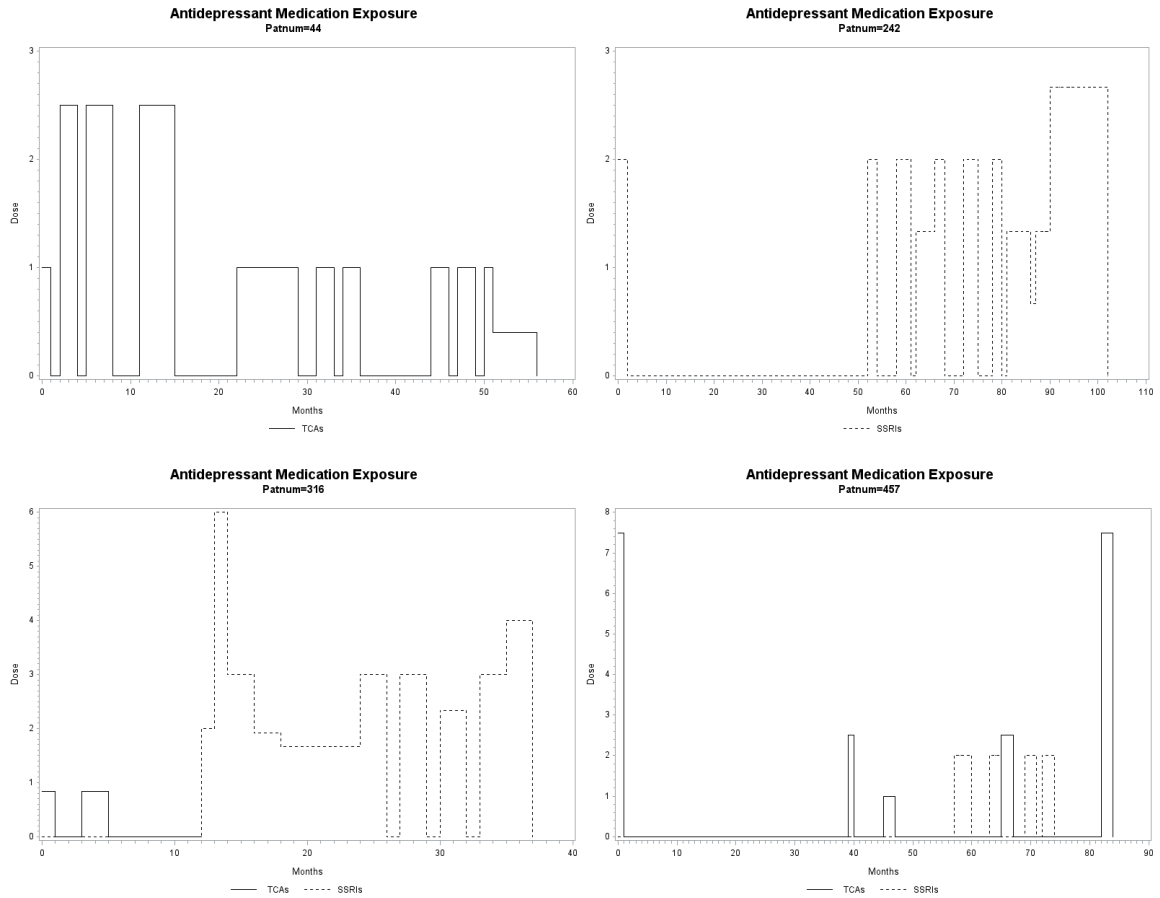


Figure 4.1: Examples of medication exposure from four typical patients. Upper left: patient who took TCA exclusively; Upper right: patient who take SSRI exclusively; Lower left: patient who switched from TCA to SSRI; Lower right: patient who took TCA and SSRI sequentially

### 4.3 Methods

#### 4.3.1 A penalized Cox PH model with time-varying exposures and interaction term

Suppose  $\{D_A(u)\}$  and  $\{D_B(u)\}$  are denoted as two series of time-varying exposures,  $A$  and  $B$ , up to time  $u$ . In practice, the time-varying exposures are usually measured at discrete time points, e.g. daily medication intake, thus that exposure history can be discretized as  $D_k(1), D_k(2), \dots, D_k(u), k = A, B$ .  $\mathbf{X}$  is an  $n \times p$  matrix for other covariates, with  $X_i$  a row vector of covariates for individual  $i$ .  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^T$  is a vector of covariate coefficients. Then the Cox Proportional Hazards (PH) model that defines the hazard at time  $u$  given the history of time-varying exposures  $\{D_k(u)\}$  and other covariates  $\mathbf{X}$  is expressed as:

$$\begin{aligned}
 h(u|\{D_A(u)\}, \{D_B(u)\}, \mathbf{X}) = & h_0(u) \exp\left(\int_0^u w_A(u-t)D_A(t)dt \right. \\
 & + \int_0^u w_B(u-t)D_B(t)dt \\
 & + \int_0^u \int_0^u w_{int}(u-t, u-s)D_A(t)D_B(s)dtds \\
 & \left. + \mathbf{X}\boldsymbol{\alpha}\right)
 \end{aligned} \tag{4.1}$$

where  $h_0(u)$  is an unspecified baseline hazard function.  $w_A(u-t)$  and  $w_B(u-s)$  are functional coefficients for time-varying exposures  $D_A(t)$  and  $D_B(s)$ . In many existing applications of exposure assessment, there was plenty of evidence for diminished effects from distant exposures. Thus, we assume that past exposure occurring certain units before has a negligible effect on current event. Equivalently the coefficient functions will smoothly go to zero at the distant boundary of the exposure interval. We impose such constraint through integral by defining the coefficient function as  $w_k(t) = \int_{-a}^{-t} s_k(c)dc, t \in [0, a_k], k = A, B$ ,

so that  $w_k(a_k) = 0$ .  $[0, a_k]$  is a supporting interval covering current time 0 and  $a_k$  time units before.  $s_k(c) = \sum_{l=1}^{L_k+4} \beta_{kl} B_{kl}(c)$ .  $L_k$  is the number of interior knots that are equally spaced on the supporting intervals. Without loss of generality, we use cubic B-spline bases,  $B_{kl}(u-t), l = 1, \dots, L_k + 4$ , and  $\beta_{kl}$  are the associated coefficients of spline bases.

The coefficient surface  $w_{int}(u-t, u-s)$  between two time-varying exposures is modeled by nonparametric bivariate function using tensor product basis as,  $w_{int}(u-t, u-s) = \sum \beta_{int,lr} B_{lr}(u-t, u-s)$ , where  $B_{lr}(u-t, u-s)$  is obtained by forming all pairwise products between the cubic B-spline basis functions as  $B_{lr}(u-t, u-s) = B_l(u-t)B_r(u-s), l = 1, \dots, L_{int} + 4, r = 1, \dots, L_{int} + 4$ , resulting a total of  $(L_{int} + 4) \times (L_{int} + 4)$  basis functions, assuming same  $L_{int}$  interior knots of both directions for simplicity. If there is only one time-varying exposure then our model is consistent to the WCE model proposed by Sylvestre and Abrahamowicz (2009).

Denote

$$\mathbf{M}_k(u) = (M_{k1}(u), \dots, M_{kL_k+4}(u)), k = A, B$$

with

$$M_{kl}(u) = \int_0^u D_k(u-t) \int_{-a_k}^{-t} B_{kl}(c) dc dt, l = 1, \dots, L_k + 4$$

Denote

$$\mathbf{N}(u) = (N_{11}(u), \dots, N_{1(L_3+4)}(u), \dots, N_{(L_3+4)(L_3+4)}(u))$$

with

$$N_{lr}(u) = \int_0^u \int_0^u B_{lr}(t, s) D_A(u-t) D_B(u-s) dt ds, l, r = 1, \dots, (L_3 + 4)$$

Let

$$\beta_k = (\beta_{k1}, \dots, \beta_{kL_k+4}), k = A, B$$

and

$$\boldsymbol{\beta}_{int} = (\beta_{int,11}, \dots, \beta_{int,1L_k+4}, \dots, \beta_{int,L_k+4L_k+4})$$

be the coefficient vectors for B-spline basis and tensor product basis. It follows that model (4.1) can be rewritten as

$$h(u) = h_0(u) \exp(\mathbf{M}_A(u)\boldsymbol{\beta}_A + \mathbf{M}_B(u)\boldsymbol{\beta}_B + \mathbf{N}(u)\boldsymbol{\beta}_{int} + \mathbf{X}\boldsymbol{\alpha}) \quad (4.2)$$

The above model using regression splines can be very useful and can be implemented using standard software. However, without any constraint on the coefficients, regression splines are more unstable and sensitive to the number and location of knots than are penalized splines. Hastie and Tibshirani (1986) gave examples where the sensitivity of regression splines to knot locations can have problems. The estimates from regression spline shows rapid, unrealistic fluctuations probably caused by substantial over-fitting of the data (Gray, 1992). The un-penalized estimation of  $\boldsymbol{\beta}$ s lead to wiggly fits and imposing constraints on  $\boldsymbol{\beta}$ s might rectify this problem. One type of constraints on estimated  $\boldsymbol{\beta}$ s is to satisfy the condition  $\boldsymbol{\beta}'_k \boldsymbol{\beta}_k < C_k, k = A, B, int$  for certain choice of  $C_k$ , which is called ridge penalty and is well known in generalized linear model. Therefore, a penalized Cox model can be fitted by applying restrictions on the coefficients of basis functions and subtracting a ridge penalty from the log likelihood, becoming the penalized partial likelihood (Perperoglou, 2014):

$$ppl(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}) = pl(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \frac{1}{2} \lambda_A \boldsymbol{\beta}_A^T \boldsymbol{\beta}_A - \frac{1}{2} \lambda_B \boldsymbol{\beta}_B^T \boldsymbol{\beta}_B - \frac{1}{2} \lambda_{int} \boldsymbol{\beta}_{int}^T \boldsymbol{\beta}_{int} \quad (4.3)$$

where  $pl(\boldsymbol{\alpha}, \boldsymbol{\beta})$  is the log partial likelihood. The factor  $\frac{1}{2}$  is introduced for mathematical convenience. The parameter estimates  $\hat{\boldsymbol{\beta}}$ s are obtained by maximizing penalized partial likelihood (4.3) when non-negative smoothing parameters  $\lambda$ s are specified. A Newton-Raphson procedure can now be used to estimate the penalized regression coefficients.

It is of crucial to determine the smoothing parameters  $\lambda$ s. If  $\lambda$ s are too small, then the  $\beta$ s are slightly shrunk to zero. If  $\lambda$ s approach infinite, the estimated  $\beta$ s would eventually become zero indicating that the estimated coefficient function would become zero lines and the estimated coefficient surface would become flat zero surface. The most common way to optimize the penalty weight is to conduct a grid search over several values of  $\lambda$ s, usually spanning from very small values to quite large values. Then the models are fitted with the different weights and the optimal choice is considered the one that maximizes some sort of criterion such as Akaike information criteria (AIC) with  $AIC = 2ppl(\alpha, \beta_1, \beta_2, \beta_3) - 2df$  (Perperoglou, 2014). In practice, it is usually more convenient to specify the degree of freedom ( $df$ ) and to solve for the values of the smoothing parameters (Buja et al., 1989). Hurvich et al. (1998) suggested a corrected AIC (AICC), which used  $n(df + 1)/(n - (df + 2))$  ( $n$  is the total number of events in Cox PH model) as the correction term in place of  $df$  in AIC, because they showed that using AIC lead to models with an excess number of degrees of freedom. We have considered both AIC and AICC in simulation studies.

The connection between smoothing parameter and degree of freedom is suggested by Gray (1992). The total degree for the model is

$$df = trace[\mathbf{I}(\boldsymbol{\eta})\mathbf{H}^{-1}(\boldsymbol{\eta}, \boldsymbol{\lambda})]$$

where  $\boldsymbol{\eta}$  is a full set of parameters  $(\alpha, \beta)$ .  $\mathbf{I}(\boldsymbol{\eta})$  is the usual observed information matrix for the partial likelihood.  $\mathbf{H}(\boldsymbol{\eta}, \boldsymbol{\lambda}) = -\partial^2 pl(\boldsymbol{\eta})/\partial \boldsymbol{\eta}^2 = \mathbf{I}(\boldsymbol{\eta}) + \mathbf{P}(\boldsymbol{\lambda})$  and  $\mathbf{P}(\boldsymbol{\lambda})$  is the second derivative matrix of the penalty function, which is a block-diagonal matrix with block  $\lambda_k \mathbf{I}_k$  corresponding to the  $\beta_k$ .  $\mathbf{I}_k$  is identity matrix of the same dimension with  $\beta_k$ .

Since the time-varying exposure  $D_k(t), k = A, B$  are usually measured at discrete time points in practical such as daily medication intake. Therefore,  $M_{kl}(u)$  can be approximated



as

$$M_{kl}(u) = \sum_0^u D_k(u-t) \left( \sum_{-a_k}^{-t} B_{kl}(c) \Delta c \right) \Delta t, l = 1, \dots, L_{k+4}$$

where  $\Delta c$  is the discrete time unit for the coefficient function and  $\Delta t$  is the discrete time unit for time-varying exposures. Similarly,  $N_{lr}(u)$  can be approximated as

$$N_{lr}(u) = \sum_0^u \sum_0^u B_{lr}(t, s) D_A(u-t) D_B(u-s) \Delta t \Delta s, l, r = 1, \dots, (L_{int} + 4)$$

where  $\Delta t$  and  $\Delta s$  are the discrete time units for time-varying exposures.

#### 4.3.2 Hypothesis test

Based on Model (4.1), one hypothesis of interest is that the effects of time-varying exposures  $D_A(t)$  and  $D_B(t)$  are additive and there is no interaction effect, leading to a flat zero coefficient surface. Thus the null hypothesis states as  $\beta_{int} = \mathbf{0}$ . It is also of interest to test whether time-varying exposure  $D_A(t)$  or  $D_B(t)$  has effect on survival outcome, so that the null hypothesis states as  $\beta_A = \mathbf{0}$  or  $\beta_B = \mathbf{0}$ .

This section discusses general linear hypotheses of the form  $\mathbf{C}\boldsymbol{\eta} = \mathbf{0}$ , where  $\mathbf{C}$  has full row rank. Following Gray (1992), a Wald type test statistic is defined as

$$(\mathbf{C}\boldsymbol{\eta})^T (\mathbf{C}\mathbf{H}^{-1}\mathbf{C}^T)^{-1} (\mathbf{C}\boldsymbol{\eta})$$

But in the penalized Cox model, this test will be conservative if a standard Chi-square statistic is assumed. Gray (1992) showed that under the null the expectation of the test statistic is approximately equal to

$$trace[(\mathbf{C}\mathbf{H}^{-1}\mathbf{C}^T)^{-1}(\mathbf{C}\mathbf{H}^{-1}\mathbf{I}\mathbf{H}^{-1}\mathbf{C}^{-T})] \quad (4.4)$$

Then, equation (4.4) will be considered as the generalized degree of freedom of the test.

## 4.4 Simulation

### 4.4.1 Simulation setup

To evaluate the performance of our proposed model, simulation studies were conducted. Three groups were considered based on types of exposure. Subjects in the first group exclusively exposed to medication  $A$ , subjects in the second group exclusively exposed to medication  $B$  and subjects in the third group exposed to both medication  $A$  and  $B$  over the follow-up period, sequentially or concurrently. Assuming equal sample size in each group, three levels of total sample size were investigated: 300, 450 and 600.

The timing of the exposure measured by days with treatment or interruption in the unit of 7 days, was generated from a lognormal distribution with mean 0.5 and standard deviation 0.8 on the log scale, i.e.  $\log(duration) \sim N(0.5, 0.8^2)$ , and rounded up to the nearest integer. The intensity of the time-varying exposure was assumed to be constant over each treatment period (in 7 day unit), which could take values of 0.005, 0.01, 0.015, 0.02, 0.025, or 0.03 during the follow-up period with equal probabilities.

Several coefficient functions with a pre-specified time window of 90 days were considered: 1) zero function ( $w(t) = 0, t \in [0, 90]$ ) indicating no exposure effect over the time window, 2) decreasing function ( $w(t) = \sin(\frac{\pi}{2} \frac{t}{90} + \frac{\pi}{2}), t \in [0, 90]$ ) indicating greater effect closer to event time with less effect for distant exposures, and U-shape function ( $w(t) = \sin(\pi \frac{t}{90} + \pi), t \in [0, 90]$ ) suggesting greater negative effect for middle exposures and less effects for recent and distant exposures. All true coefficient functions smoothly go to zero at the distant boundary of the exposure interval. The coefficient functions were defined on reversed time axes so that day zero in Figure 4.2 corresponds to the time of cumulative effects measured

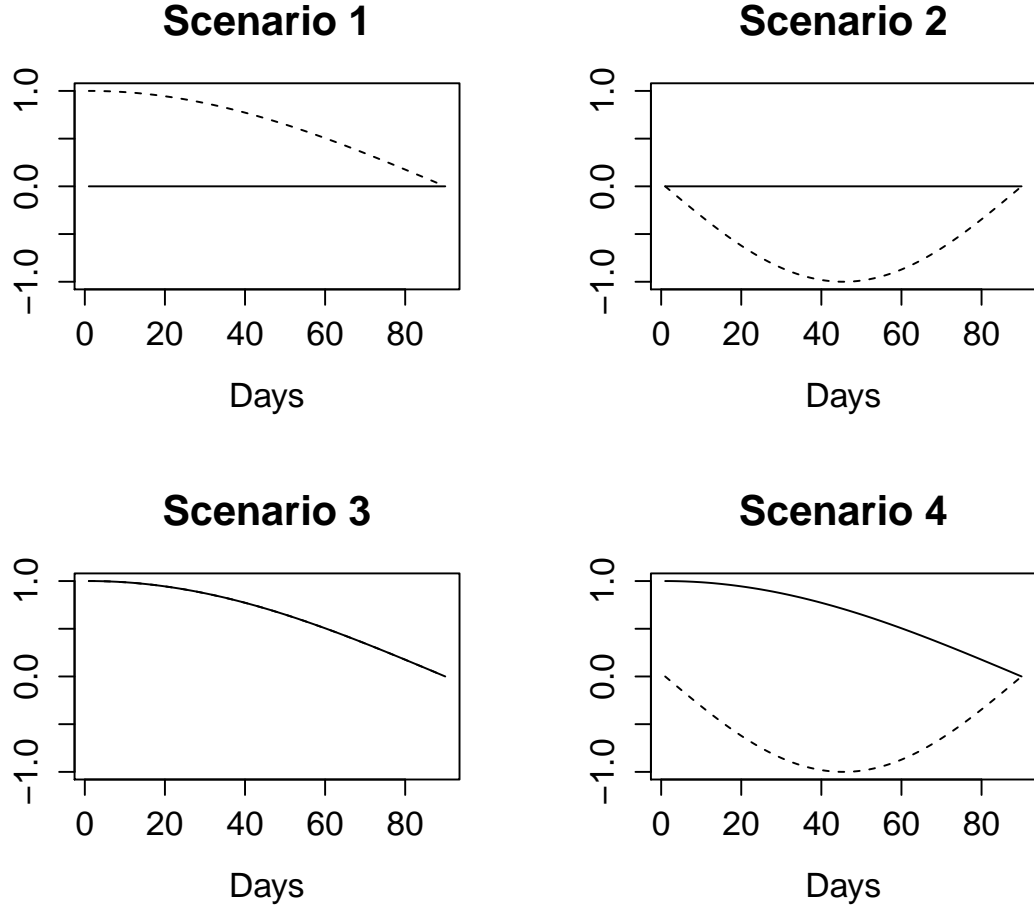


Figure 4.2: Four scenarios of coefficient functions used in simulations:  $w_{(A)}(t)$  is indicated by solid lines;  $w_{(B)}(t)$  is indicated by dashed lines. The coefficient functions were defined on reversed time axes so that day zero corresponds to the time of outcome measures and increasing days correspond to more distant past. Scenario 3 has identical coefficient functions.

and increasing days correspond to more distant past. We generated 200 data sets under each scenario as shown in Figure 4.2.

Coefficient surface with a pre-specified time window of 30 days were generated from two individual surfaces as shown in Figure 4.3. The interaction surface has higher values along the diagonal, indicating that medications have larger interaction effect when taken simultaneously and such effect decreases while approaching to the distant boundary. Similar to the definition of coefficient function, the interaction surface was also defined on reversed

time axes. X-axis indicates the exposure time for medication A and y-axis represents medication B.

Survival outcomes were generated based on the following formula with various combination of corresponding coefficient functions:

$$\begin{aligned}
h(u|\{D_A(u)\}, \{D_B(u)\}) = & h_0(u) \exp\left(\int_0^u w_A(u-t)D_A(t)dt \right. \\
& + \int_0^u w_B(u-t)D_B(t)dt \\
& \left. + \int_0^u \int_0^u w_{int}(u-t, u-s)D_A(t)D_B(s)dtds\right) \quad (4.5)
\end{aligned}$$

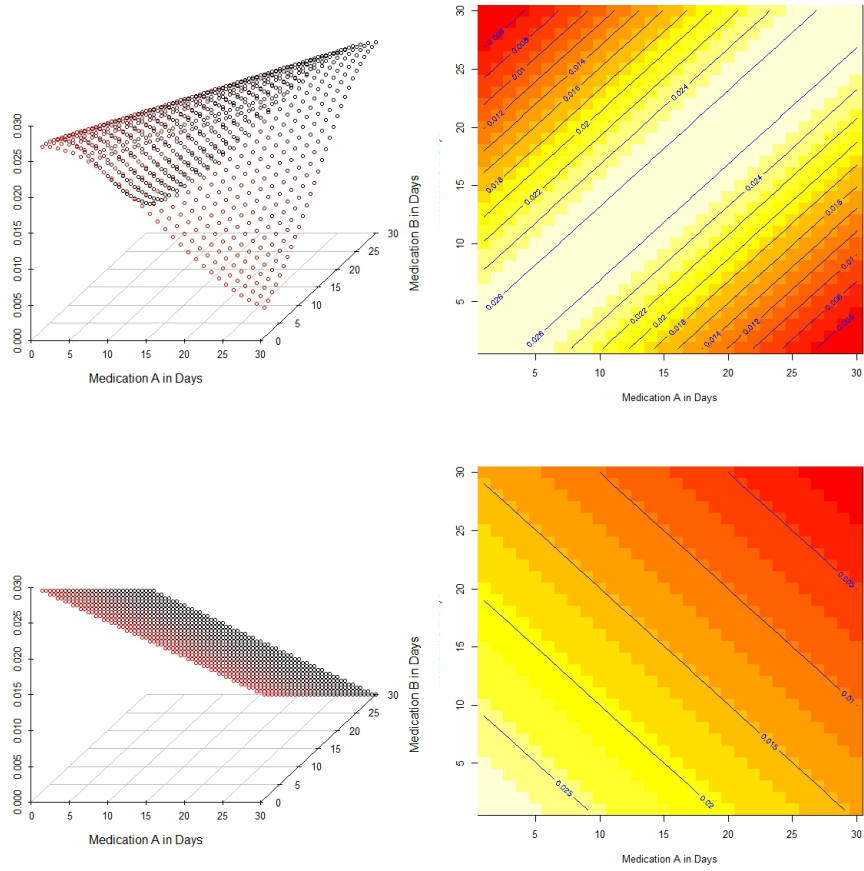
Permutation algorithm was used, which is designed and validated for simulating survival times with time-varying covariates Sylvestre and Abrahamowicz (2008).

#### 4.4.2 Simulation results

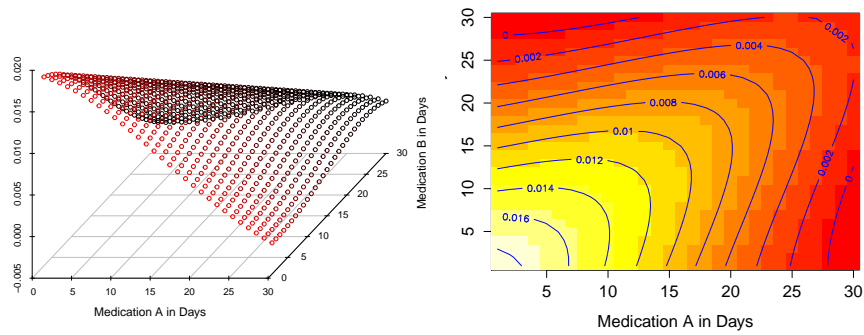
We simulated 200 data sets and the proposed Model (4.1) was fitted with six equally spaced interior knots for each coefficient function and one interior knots for coefficient surface. AIC and AICC were both used to select smoothing parameters. Since we adopted cubic B-spline basis, the maximum degrees of freedom for coefficient function and coefficient surface are 10 and 25 respectively.

Estimates  $\hat{w}_A$ ,  $\hat{w}_B$  and  $\hat{w}_{int}$  were obtained for each scenario. The average mean squared error for the interaction is calculated as

$$avgMSE_{int} = \frac{1}{200} \sum_{r=1}^{200} \frac{1}{30 \times 30} \sum_{t=1}^{30} \sum_{s=1}^{30} (w_{int}(t, s) - \hat{w}_{int}(t, s))^2 \quad (4.6)$$



(a) Two surfaces with associated contour plots used to generate interaction surface



(b) Interaction surface (3D)

(c) Interaction surface (Contour)

Figure 4.3: Two 3D surfaces with associated contour plots used to generate the interaction surface with a pre-specified time window of 30 days. Surfaces were defined on reversed time axes, where x-axis indicates the exposure time for medication A and y-axis represents medication B. Upper surface has constant highest values on the diagonal. Middle surface has decreasing values with increasing days. Lower surface is a weighted combination of upper surface and middle surface.

where the index  $r$  represents the replicates.  $avgMSE_A$  and  $avgMSE_B$  are defined analogously as

$$avgMSE_k = \frac{1}{200} \sum_{r=1}^{200} \frac{1}{90} \sum_{t=1}^{90} (w_k(t) - \hat{w}_k(t))^2, \quad k = A, B \quad (4.7)$$

Results of  $avgMSE$  are presented in Table 4.1. The variability of estimated coefficient functions and coefficient surface decreases as sample size increases. Models that use corrected AIC (AICC) to select smoothing parameter tend to have slightly better performance compared with models using AIC. Also, the original shape of coefficient function influences the estimation in terms of  $avgMSE$ .

Figure 4.4 to Figure 4.15 show the estimated coefficient functions and mean of estimated coefficient surface from Scenario 1 to Scenario 4 with sample size of 300, 450 and 600. In each panel of coefficient function, the true coefficient functions are denoted as white solid lines and 200 estimates as black lines. Less fluctuation was observed as sample size increases. The estimated functions or surface were able to capture the shape of the true ones. Figure 4.16 and Figure 4.16 show the sample-based mean and point-wise confidence bands across 200 replications. The 2.5% and 97.5% point-wise quantiles are denoted as red dashed lines. Mean estimates are denoted as blue dashed lines and the true coefficient functions used to simulate data are denoted as black solid lines.

Table 4.2 presents the empirical type I errors of hypothesis tests under sample size of 450 over 200 replications. Two hypothesis tests of interest were conducted. One is to test whether there is an interaction between two time-varying exposures. Under the null hypothesis of no interaction, the coefficient surface is expected to be a zero flat surface. Wald test in penalized Cox's PH model discussed in Section 4.3.2 was performed by testing

Table 4.1: Average mean squared error (avgMSE) ( $\times 10^{-5}$ ) of coefficient functions and coefficient surface estimated by penalized Cox PH method over 200 replications

Coefficient	Sample size	Selection method	Scenarios			
			S1	S2	S3	S4
$w_A(t)$	300	(AIC)	4.40	4.76	9.11	9.06
	300	(AICC)	3.58	3.86	8.84	8.26
	450	(AIC)	3.08	3.66	5.72	5.43
	450	(AICC)	2.96	3.33	5.44	5.24
	600	(AIC)	2.26	2.30	5.05	4.46
	600	(AICC)	2.16	2.14	4.77	4.07
$w_B(t)$	300	(AIC)	7.91	12.4	7.97	14.7
	300	(AICC)	6.93	12.1	7.59	14.0
	450	(AIC)	5.60	9.60	5.82	9.39
	450	(AICC)	5.31	9.41	5.68	9.15
	600	(AIC)	4.44	8.08	4.67	7.71
	600	(AICC)	4.37	8.07	4.50	7.61
$w_{int}(t)$	300	(AIC)	2.45	2.28	2.37	1.84
	300	(AICC)	1.83	1.89	2.00	1.52
	450	(AIC)	1.29	1.20	1.88	1.17
	450	(AICC)	1.19	1.08	1.68	1.07
	600	(AIC)	0.81	0.93	1.24	0.88
	600	(AICC)	0.77	0.89	1.16	0.80

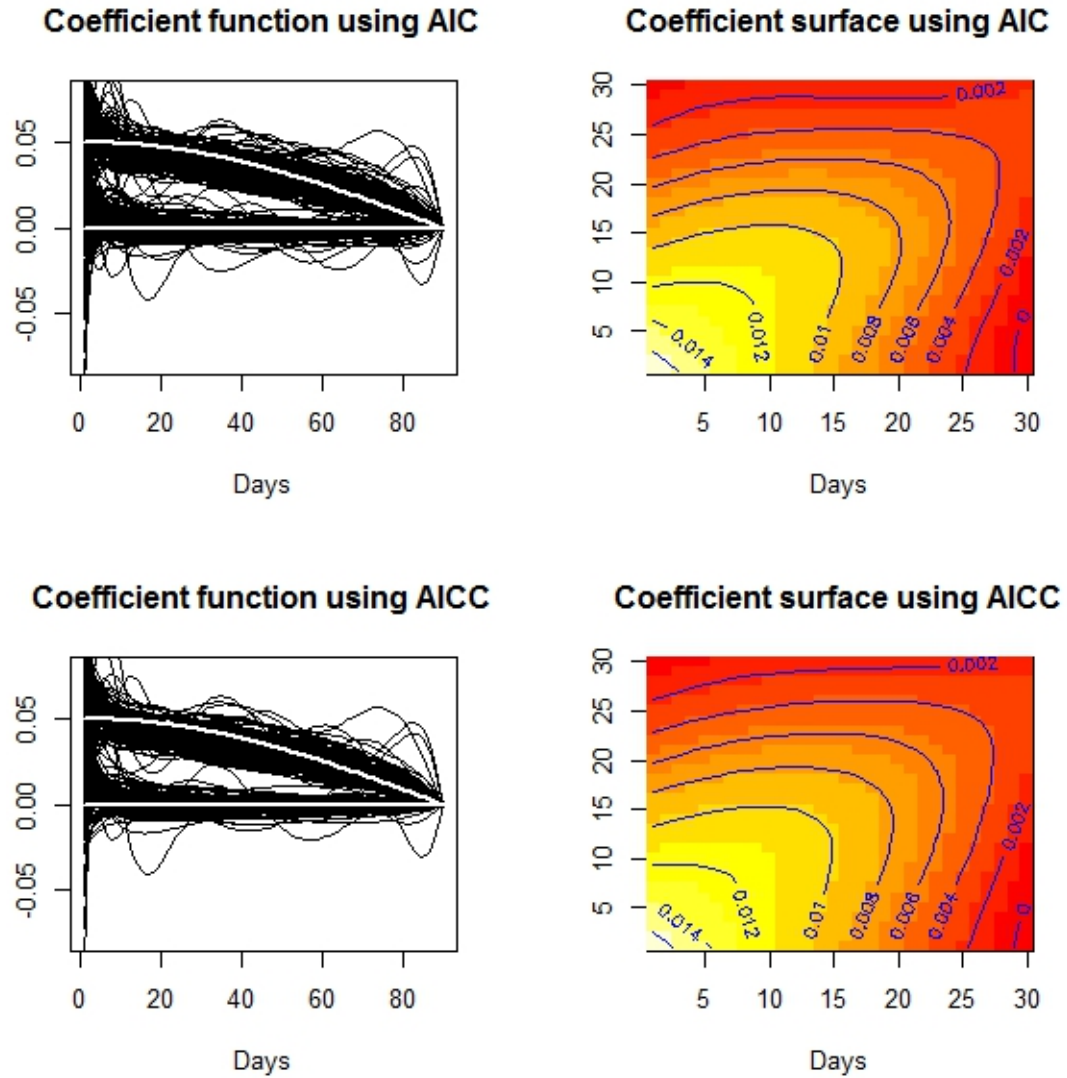


Figure 4.4: Estimated coefficient functions and mean of estimated coefficient surface in Scenario 1 with sample size of 300. True coefficient functions are denoted as white solid lines.



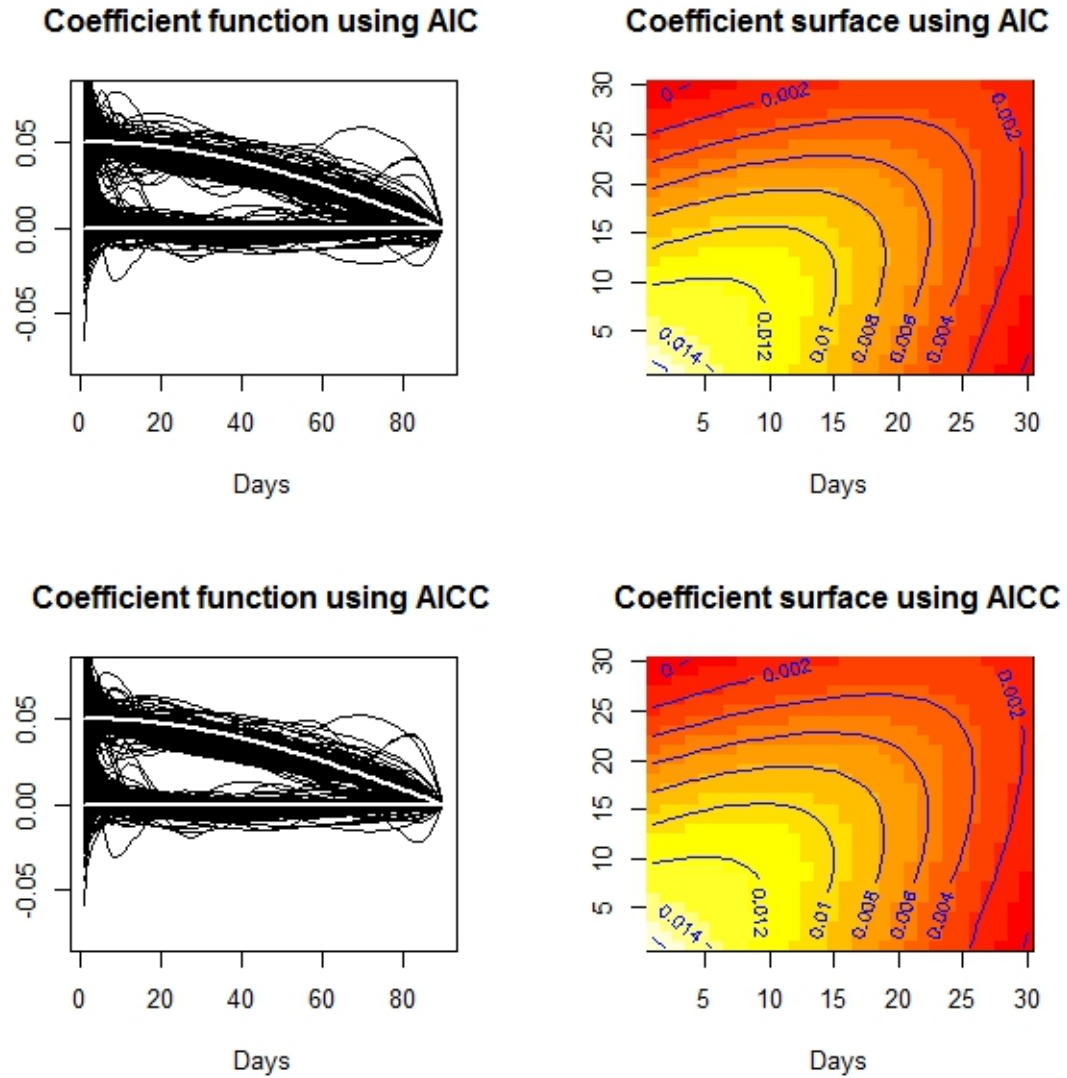


Figure 4.5: Estimated coefficient functions and mean of estimated coefficient surface in Scenario 1 with sample size of 450. True coefficient functions are denoted as white solid lines.

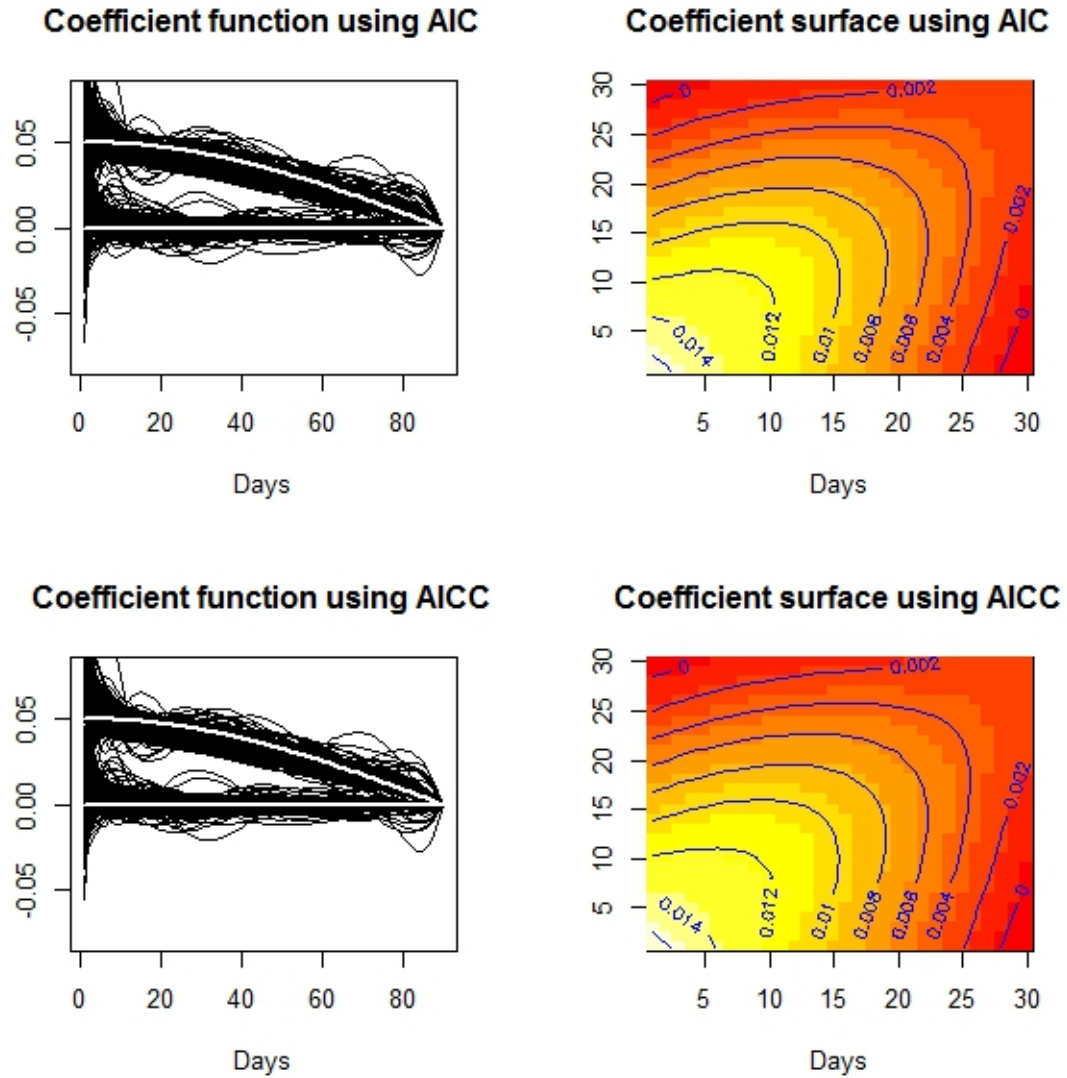


Figure 4.6: Estimated coefficient functions and mean of estimated coefficient surface in Scenario 1 with sample size of 600. True coefficient functions are denoted as white solid lines.

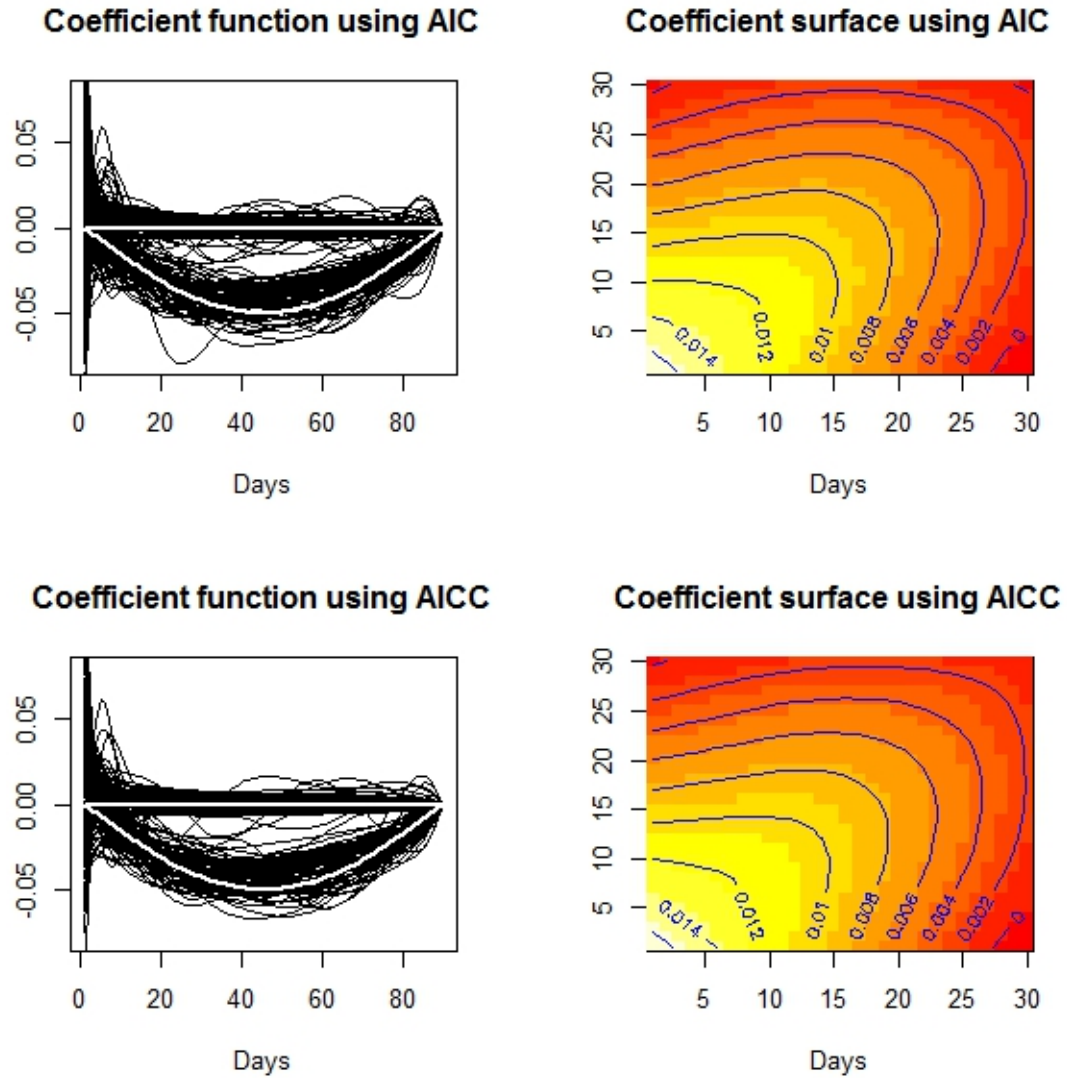


Figure 4.7: Estimated coefficient functions and mean of estimated coefficient surface in Scenario 2 with sample size of 300. True coefficient functions are denoted as white solid lines.

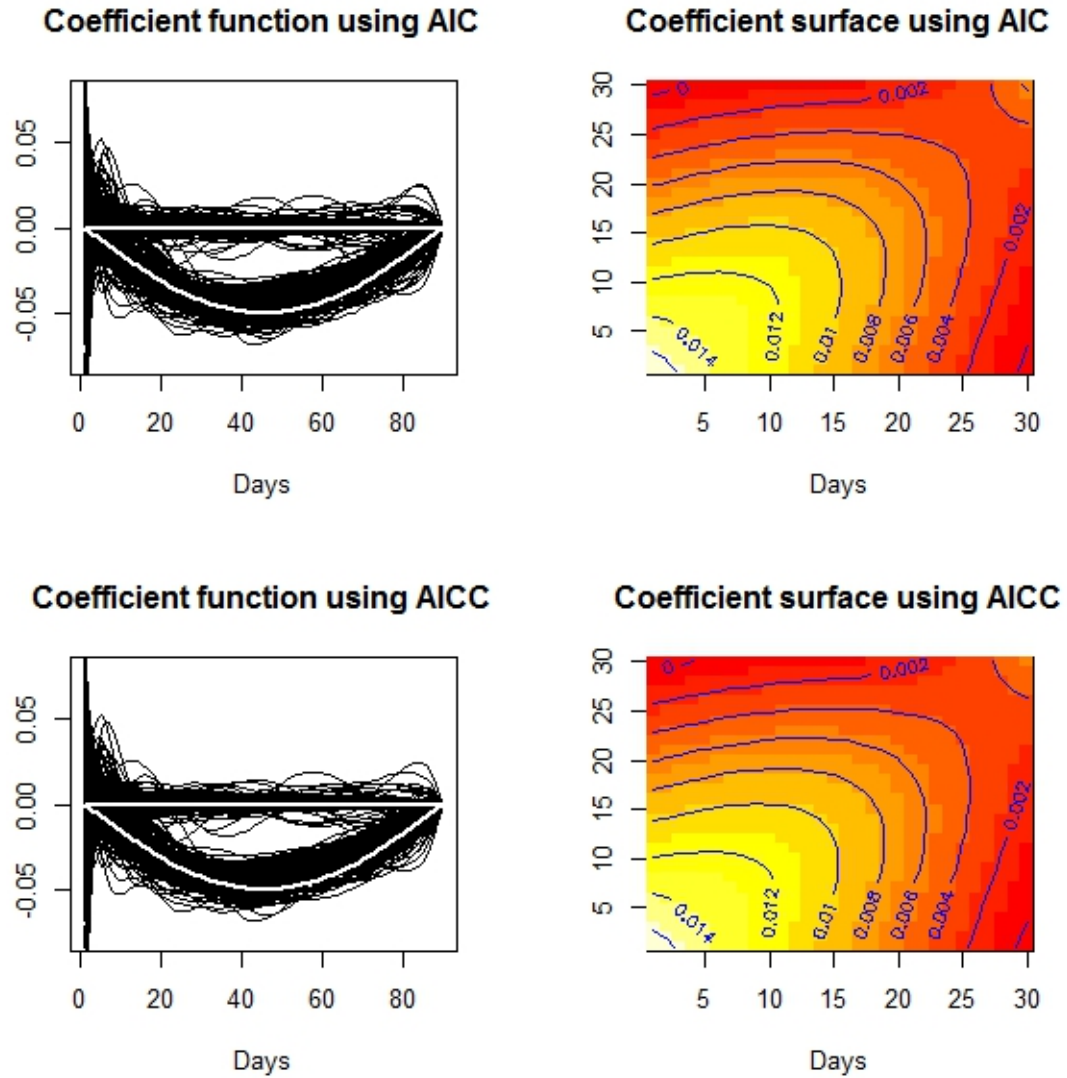


Figure 4.8: Estimated coefficient functions and mean of estimated coefficient surface in Scenario 2 with sample size of 450. True coefficient functions are denoted as white solid lines.

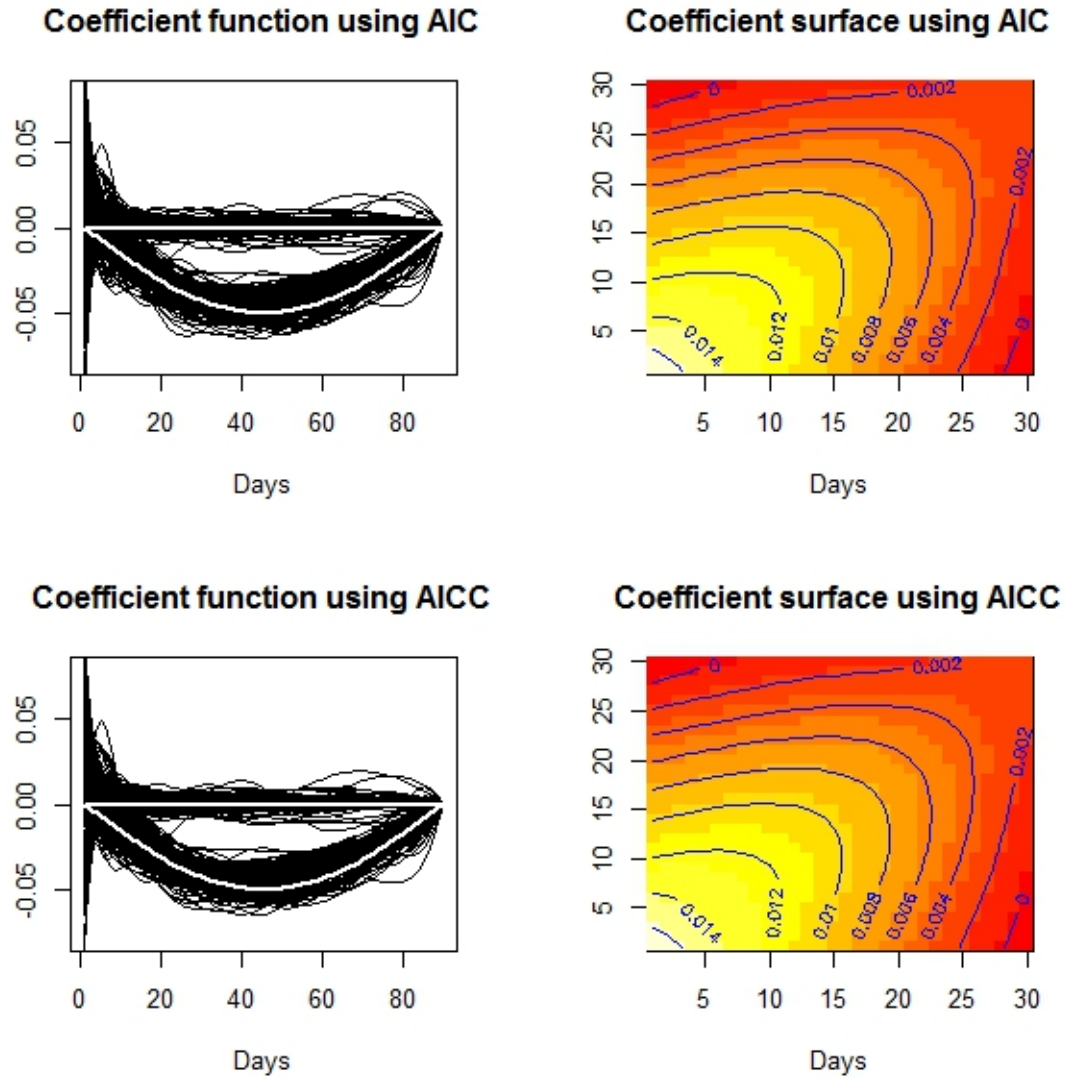


Figure 4.9: Estimated coefficient functions and mean of estimated coefficient surface in Scenario 2 with sample size of 600. True coefficient functions are denoted as white solid lines.

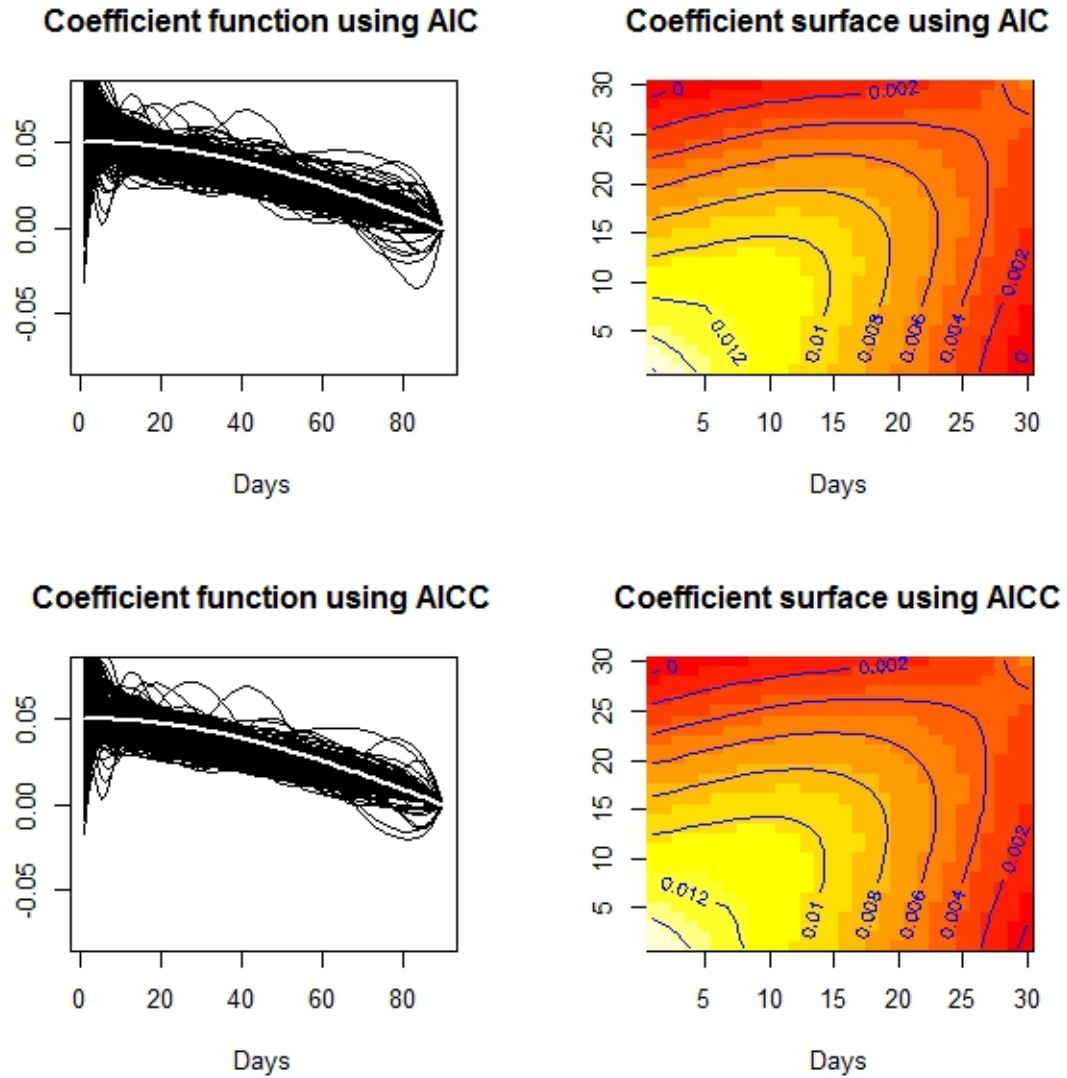


Figure 4.10: Estimated coefficient functions and mean of estimated coefficient surface in Scenario 3 with sample size of 300. True coefficient functions are denoted as white solid lines.

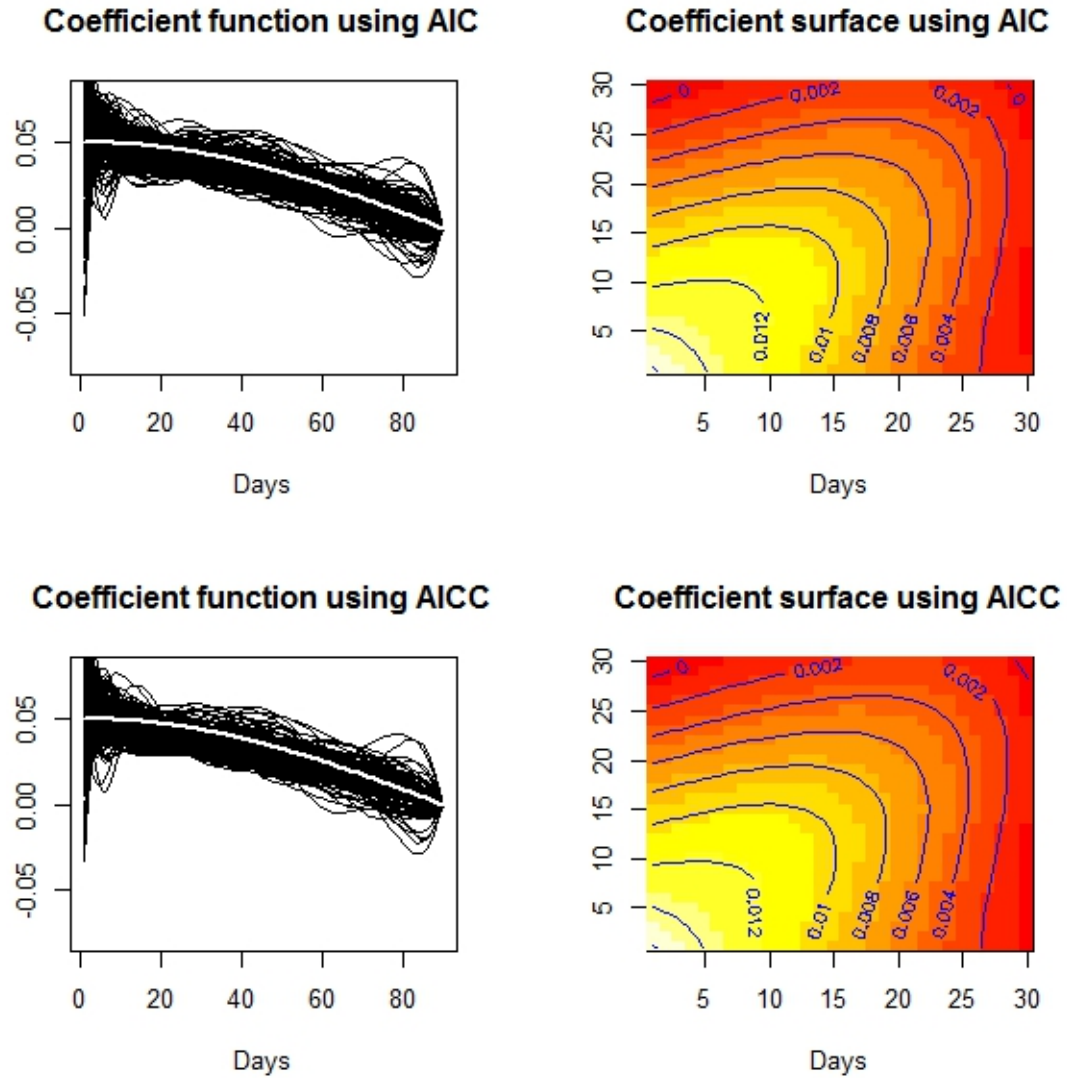


Figure 4.11: Estimated coefficient functions and mean of estimated coefficient surface in Scenario 3 with sample size of 450. True coefficient functions are denoted as white solid lines.



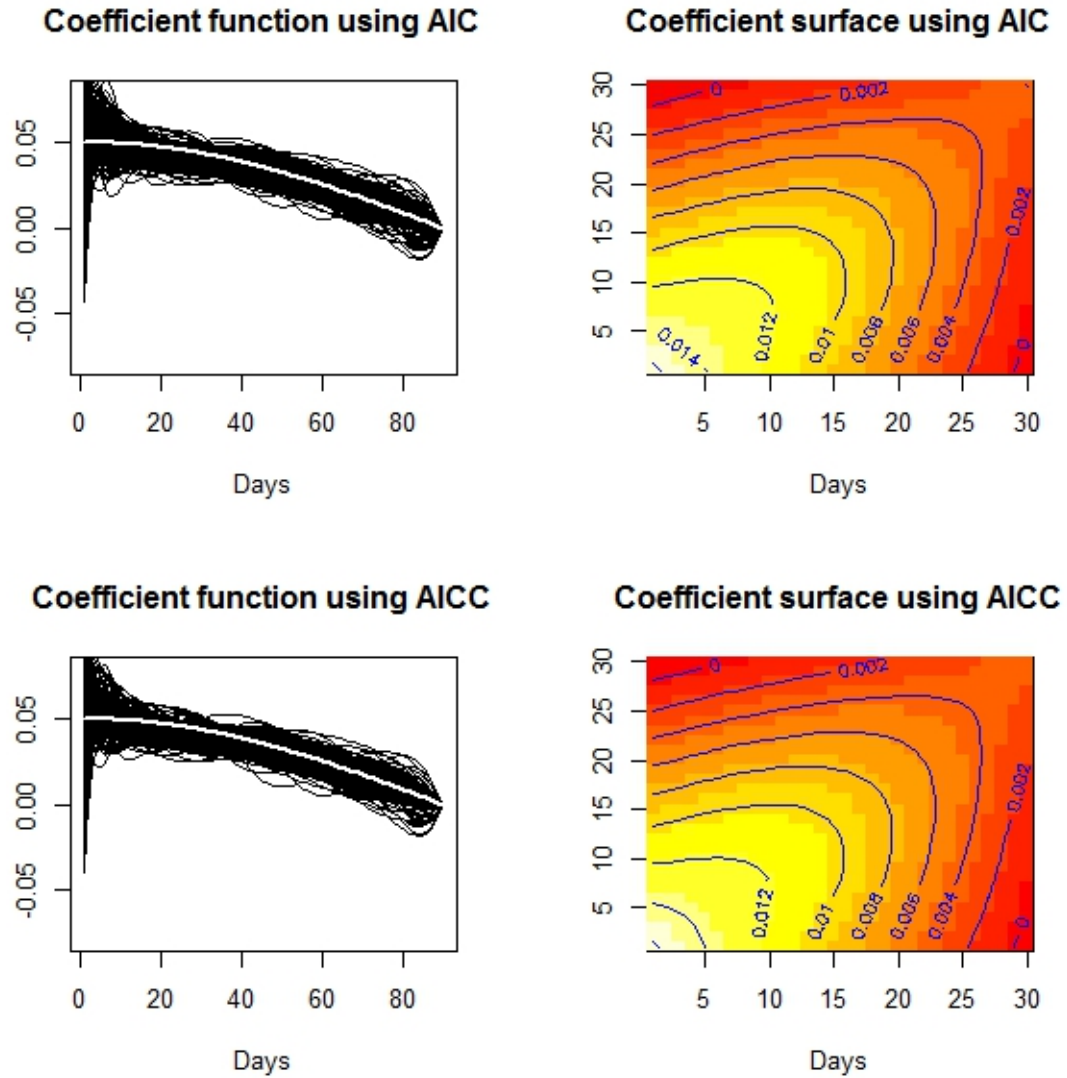


Figure 4.12: Estimated coefficient functions and mean of estimated coefficient surface in Scenario 3 with sample size of 600. True coefficient functions are denoted as white solid lines.



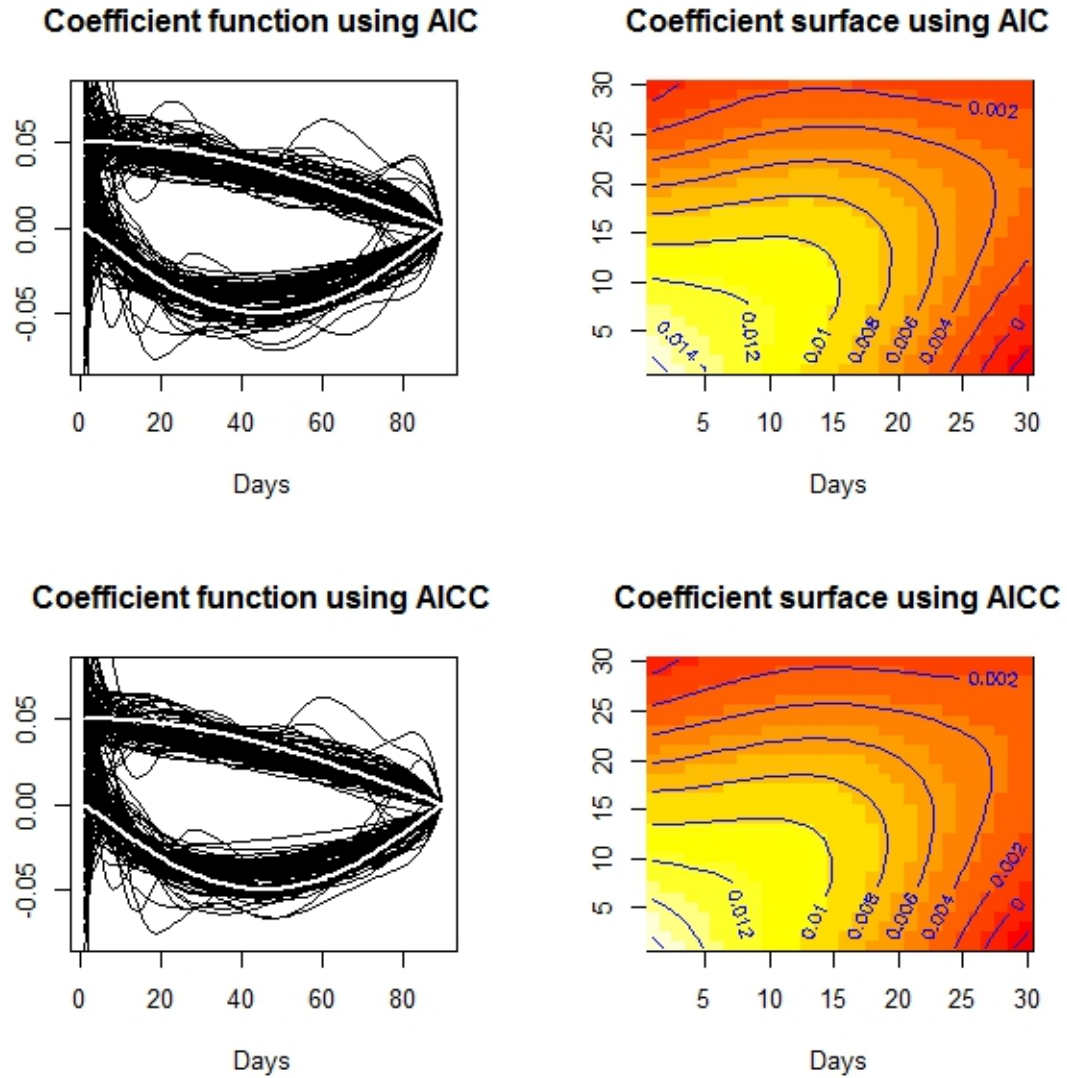


Figure 4.13: Estimated coefficient functions and mean of estimated coefficient surface in Scenario 4 with sample size of 300. True coefficient functions are denoted as white solid lines.

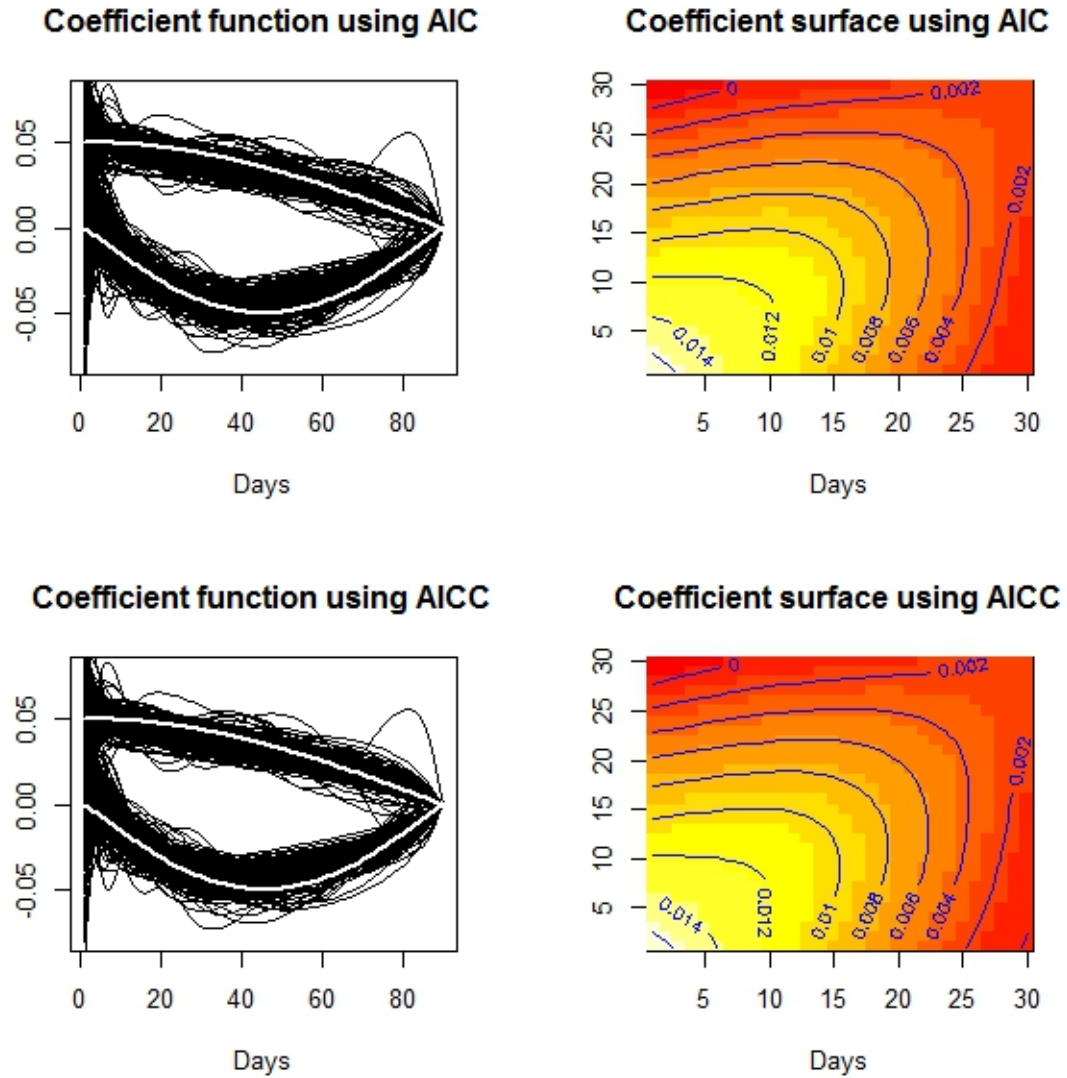


Figure 4.14: Estimated coefficient functions and mean of estimated coefficient surface in Scenario 4 with sample size of 450. True coefficient functions are denoted as white solid lines.

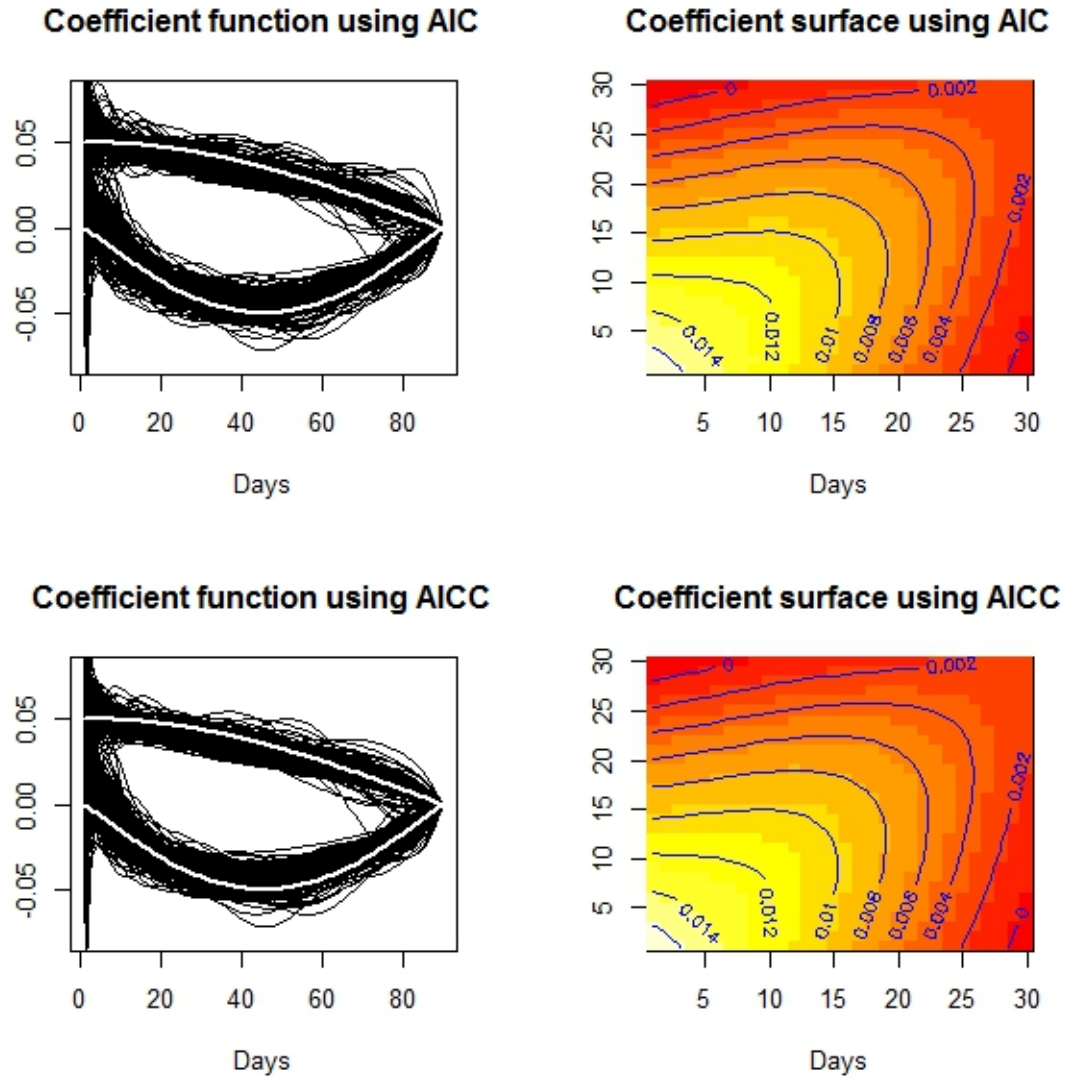
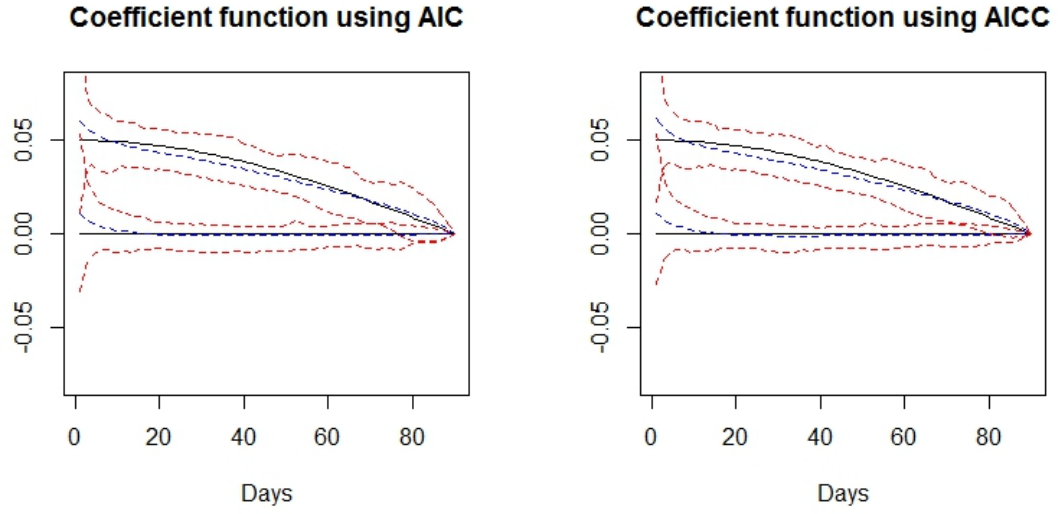
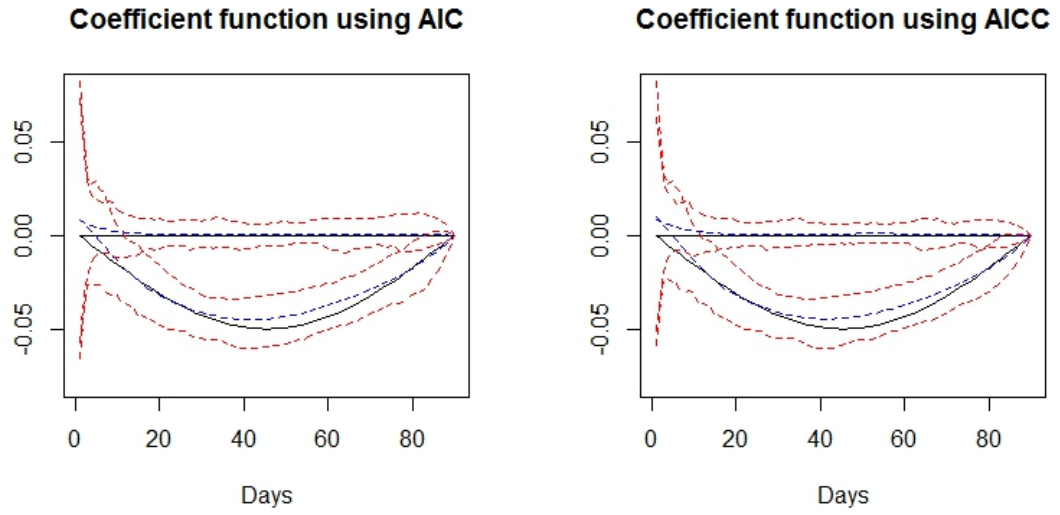


Figure 4.15: Estimated coefficient functions and mean of estimated coefficient surface in Scenario 4 with sample size of 600. True coefficient functions are denoted as white solid lines.

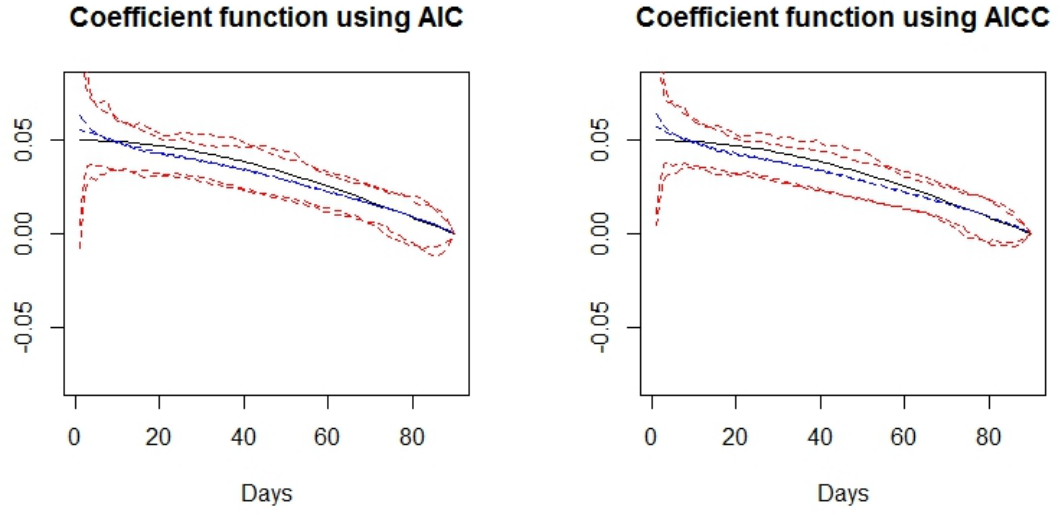


(a) Scenario 1

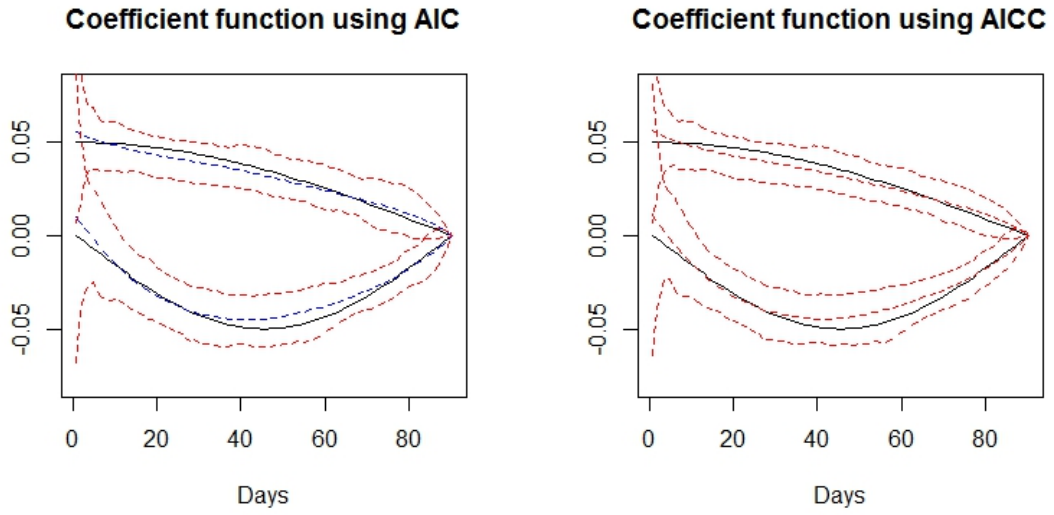


(b) Scenario 2

Figure 4.16: Sample point-wise confidence bands for Scenario 1 and Scenario 2. Red dashed lines are 2.5% and 97.5% point-wise quantiles. Blue dashed lines are mean estimates. The true coefficient functions are denoted as black solid lines.



(a) Scenario 3



(b) Scenario 4

Figure 4.17: Sample point-wise confidence bands for Scenario 3 and Scenario 4. Red dashed lines are 2.5% and 97.5% point-wise quantiles. Blue dashed lines are mean estimates. The true coefficient functions are denoted as black solid lines.

Table 4.2: Empirical type I errors of Wald tests under sample size of 450 across 200 replications

Null hypothesis (Selection method)	Scenarios			
	S1	S2	S3	S4
No interaction (AIC)	0.06	0.07	0.05	0.03
No interaction (AICC)	0.05	0.07	0.05	0.03
No effect of $D_{(A)}(t)$ (AIC)	0.06	0.04	-	-
No effect of $D_{(A)}(t)$ (AICC)	0.06	0.04	-	-

$\beta_{int,lr} = 0, l, r = 1, \dots, 5$ . In Scenario 1 and Scenario 2,  $w_{AS}$  are flat zero lines indicating no main effect from exposure  $A$ . Therefore, hypothesis tests for the main effect of  $A$  were also performed under Scenario 1 and Scenario 2 by testing  $\beta_{Al} = 0, l = 1, \dots, 10$ . Table 4.2 shows that all the empirical type I errors were close to the nominal level 0.05 and selection methods using AIC or AICC have comparable rate.

#### 4.5 Application

In this section we revisit the motivating data set we described in Section 2 and apply the proposed model. Our primary interest is to examine the effects of two classes of antidepressants on the risk of coronary artery disease (CAD). The analysis data set contained 297 subjects in total with 160 patients taking TCAs exclusively, 93 patients taking SSRIs exclusively and 44 patients taking both. In this elderly population, patients were also at risk of dying. Patients who died of myocardial infarction (MI) without previous CAD are also considered as CAD events. In order to combine such information, death dates were extracted from Medicare/Medicaid data set and causes of death were obtained from death certificates provided by the Indiana State Department of Health. Therefore, our primary outcome, a hard CAD event, was defined as the occurrence of any of the following events in the medical record or Medicare/Medicaid data during the follow-up period: (a) fatal

MI; (b) laboratory evidence of acute MI (Creatine kinase-myocardial band isoenzyme value  $> 3.0$  ng/ml or troponin value  $> 0.3$  ug/L); and (c) diagnosis of CAD.

Entry time in the survival analysis was the first time when a patient was dispensed an antidepressant (Suissa, 2007). Event time was calculated using CAD date if a patient had CAD or death date if a patient died of CAD. Censoring time was calculated using last clinical encounter date for those who were alive or death data if a patient died of other disease. A total of 91 patients (30%) developed hard CAD during the follow-up.

In the model, some potential covariates such as gender and age at first serve of medication were adjusted. Table 4.3 shows descriptive statistics among three groups. According to previous simulation result, models with selection method of corrected AIC (AICC) have better estimation results, thus we applied models with AICC and estimated the coefficient functions and coefficient surface under various pre-specified exposure time windows using cubic B-spline basis with 10 basis functions and cubic tensor product basis with 25 basis surfaces. Table 4.4 summarizes estimates of some coefficients and p-values obtained from hypothesis tests under various pre-specified relevant time (in month). The length of pre-specified time interval for interaction surface measures the longest drug-drug interaction interval. For example, 18-month time interval for coefficient surface indicates that no interaction effect between SSRI at current time and TCA at 18 months before, or vice versa. By assuming that influence time for drug-drug interaction is not longer than that for medication itself, we only present results where time interval of coefficient function is equal or longer than the interval of coefficient surface. Regardless of time interval, results from the Wald tests for all the models indicate that there is no interaction surface.

Table 4.5 shows under various time intervals, estimates of female and age at first serve of antidepressants and P-values of hypothesis tests of time-varying antidepressants effect on risk of CAD from models given no interaction term. Antidepressant class of TCAs shows

Table 4.3: Comparison of demographic information among three groups. P-value is calculated using ANOVA for continuous variables and Chi-square test for binary variables

	SSRI (n=93)	TCA (n=160)	TCA&SSRI (n=44)	P-value
Age at first serve Mean(SD)	73.62(6.15)	70.21(6.88)	69.73(6.39)	0.0001
Years of Education Mean(SD)	8.91(2.67)	8.76(2.83)	8.48(3.45)	0.7094
Female (%)	80(86.02)	120(75)	35(79.55)	0.1147
African-American (%)	69(74.19)	115(71.88)	24(54.55)	0.0484
Smoke at any time (%)	31(33.33)	69(43.13)	16(36.36)	0.2827
CAD Events (%)	27(29.03)	53(33.13)	11(25)	0.5389

marginally significance. Figure 4.18 shows estimated coefficient functions for both TCAs and SSRIs under various pre-specified time interval. Black lines indicate TCAs. Red lines indicate SSRIs. The timing-varying effect of TCAs can be explained as TCAs is marginally associated with the risk of CAD event and TCAs taken recently have larger influence than that taken more distant before.

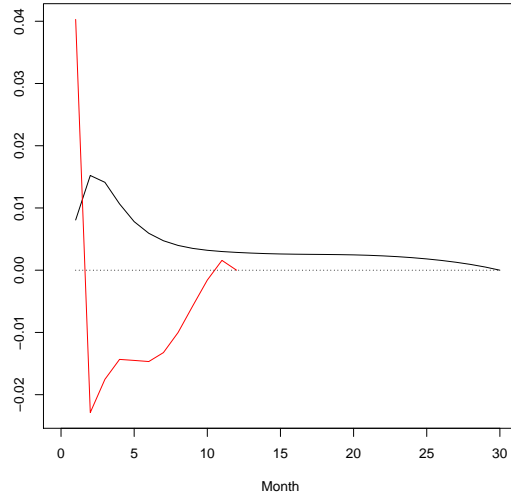


Table 4.4: Estimates and P-values of hypothesis test of no interaction effect under various pre-specified time interval

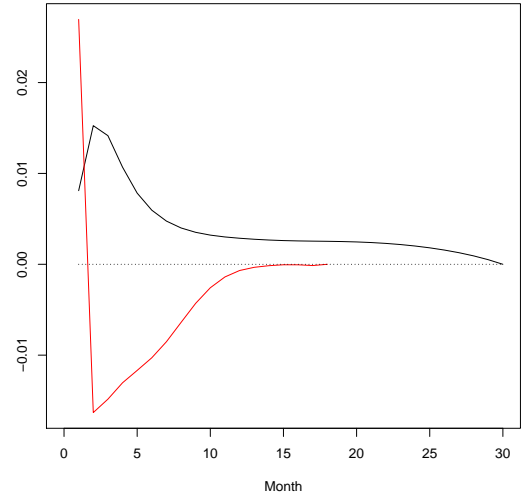
Time interval(in month)		Covariates Mean(SD)		P-values of hypothesis test of no interaction effect
coefficient function	coefficient surface	Female	Age	
12	12	-0.33(0.27)	0.05(0.02)	0.97
18	12	-0.33(0.27)	0.05(0.02)	0.97
18	18	-0.33(0.27)	0.05(0.02)	0.94
24	12	-0.33(0.27)	0.05(0.02)	0.97
24	18	-0.33(0.27)	0.05(0.02)	0.94
24	24	-0.33(0.27)	0.05(0.02)	0.82
30	12	-0.33(0.27)	0.05(0.02)	0.97
30	18	-0.32(0.27)	0.04(0.02)	0.94
30	24	-0.32(0.27)	0.05(0.02)	0.83
30	30	-0.32(0.27)	0.04(0.02)	0.88

Table 4.5: Estimates and P-values of hypothesis test of no interaction effect under various pre-specified time interval

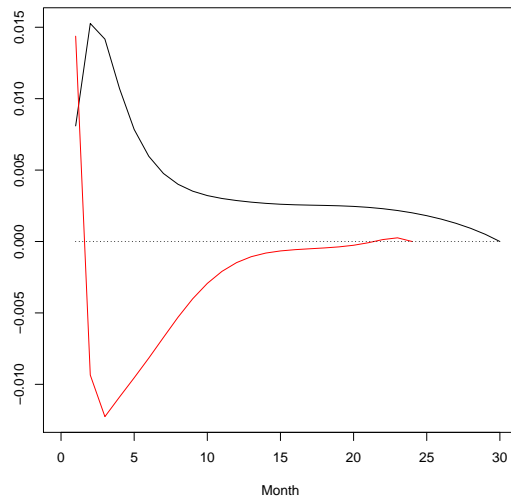
Time interval of coefficient function (in month)		Covariates Mean(SD)		Test of no time-varying effect (P-value)	
TCA	SSRI	Female	Age	TCA	SSRI
12	12	-0.33(0.27)	0.05(0.02)	0.07	0.67
12	18	-0.33(0.27)	0.05(0.02)	0.07	0.75
12	24	-0.33(0.27)	0.05(0.02)	0.07	0.80
12	30	-0.33(0.27)	0.05(0.02)	0.07	0.76
18	12	-0.33(0.27)	0.05(0.02)	0.09	0.41
18	18	-0.33(0.27)	0.05(0.02)	0.09	0.75
18	24	-0.33(0.27)	0.05(0.02)	0.09	0.81
18	30	-0.33(0.27)	0.05(0.02)	0.09	0.76
24	12	-0.33(0.27)	0.05(0.02)	0.09	0.41
24	18	-0.33(0.27)	0.05(0.02)	0.09	0.75
24	24	-0.33(0.27)	0.05(0.02)	0.09	0.81
24	30	-0.33(0.27)	0.05(0.02)	0.10	0.76
30	12	-0.33(0.27)	0.05(0.02)	0.09	0.41
30	18	-0.33(0.27)	0.05(0.02)	0.09	0.75
30	24	-0.33(0.27)	0.05(0.02)	0.09	0.81
30	30	-0.33(0.27)	0.05(0.02)	0.10	0.76



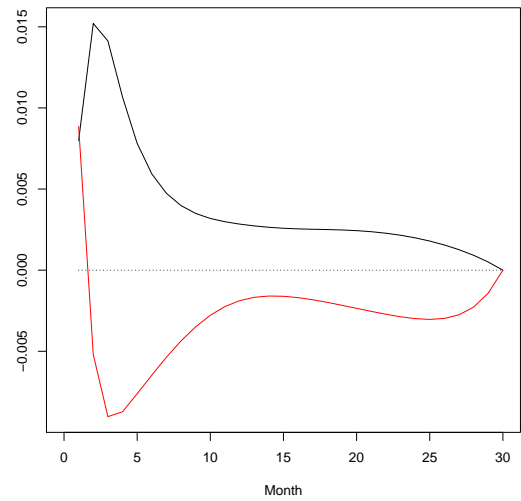
(a) 30 month TCA vs 12 month SSRI



(b) 30 month TCA vs 18 month SSRI



(c) 30 month TCA vs 24 month SSRI



(d) 30 month TCA vs 30 month SSRI

Figure 4.18: Estimates of coefficient functions for both TCA and SSRI under various pre-specified time interval. Black lines indicate TCA. Red lines indicate SSRI

## 4.6 Discussion

In this chapter, we proposed a penalized Cox proportional hazards (PH) model with functional covariates and interaction term, which can be used to examine the effects of time-varying exposures and their interaction effect on survival outcome. Flexible weighted cumulative exposure (WCE) model proposed by Sylvestre and Abrahamowicz (2009) and penalized scalar-on-functions regression model by Fuchs et al. (2015) provide insight into the formula of our model. Penalized Cox PH model discussed by Perperoglou (2014) and Gray (1992) provide methods of parameter estimation. Furthermore, parameter estimation can be implemented using standard statistical packages for survival analysis with time-dependent covariates making the proposed models applicable to many data applications (Therneau et al., 2003). Simulation studies demonstrate adequate performance of the proposed methods in parameter estimation as well as hypothesis testing. The application of the proposed method in a cohort of depressed patients shows that there is no interaction between two classes of antidepressants. The proposed model framework can be readily applied to other medication groups into determine whether medication use is associated with other survival health outcomes.

The models considered here have constraint on the right end of coefficient functions for biological and identification reason. In the flexible WCE model of single exposure, Sylvestre and Abrahamowicz (2009) imposed such constraint by forcing the last two coefficients of basis functions to be zero, leading to the loss of degree of freedom. Compared with their method, the way we put the constraint doesn't have such issue. Another strength is that the units between two types of exposure are not necessarily to be converted to the same magnitude. In our model, SSRIs and TCAs are two classes of antidepressants. Within each class standard dosage was derived among antidepressants because they were considered as one type of exposure, but it is unnecessary to adjust dosage between TCAs and SSRIs.

Another significant development based on our methods is to model interactions among the different medications patients are taking. In aging studies involving patients with multiple chronic conditions, the majority of patients were found to be on multiple medications at any given period. Thus it is important that statistical models take into account all medications a patient has been exposed in order to detect potential risk to adverse events.

Therneau et al. (2003) discussed the natural connection between penalized Cox PH model and frailty model. One possible extension of our model is that the penalized parameters can be considered as random effects. In our model, we did a grid search to select smoothing parameter based on AIC and AICC. Thus, our model is computational demanding and highly depends on the intensity of grid. By assuming penalized parameters from standard Gaussian distribution, the model may have advantage of using frailty model methodology to optimize the penalty weight.

In summary, we proposed a penalized Cox PH model with two time-varying exposures and their interaction. The proposed methods performed adequately in simulation studies. With the increasing availability of electronic medical records of medication dispensing data, our proposed methods can be readily applied to other medication dispensing data to determine whether multiple medication use have a long-term association with survival health outcomes.

## Chapter 5

### Conclusion

We have studied several flexible models related to time-varying exposures. The proposed methodologies are applicable to many medical research areas.

First, in Chapter 2, we developed functional regression models for multivariate outcomes with time-varying exposures using penalized splines. Our simulation studies indicate adequate performances for parameter estimation and inferences. Our proposed model extended previous methods in several ways. First, this model included multivariate longitudinally measured outcomes and estimate the effects of exposure on all the outcomes simultaneously. Second, our model framework included a constraint at the boundary of the exposure interval and hence reduces the estimation variability. By expressing the functional model into a mixed model representation, we showed that parameter estimation and inference procedures can be implemented using standard statistical software packages. We also demonstrated the application of the proposed method in a group of SSRIs users in estimating the association between SSRIs exposure and longitudinal blood pressure measures. The proposed model framework can be readily applied to other medication groups to determine whether medication use is associated with longitudinally measured health outcomes.

Second, in Chapter 3, we extended the flexible WCE model proposed by Sylvestre and Abrahamowicz (2009) for single exposure to compare multiple exposures. The use of flexible weight functions using spline bases not only offered flexibility in summarizing cumulative exposures, but also provided information on how the risk of adverse event was impacted by the timing of the exposures. Furthermore, parameter estimation for the proposed model can be implemented using standard statistical packages for survival analysis with time-

dependent covariates making the proposed models applicable to many data applications. Our simulation studies demonstrated adequate performance of the proposed methods in parameter estimation as well as hypothesis testing using fixed number of knots. However, simulation results also revealed inflated Type I error rates when the test was conditional on AIC criteria selected number of knots. Further research is needed to determine a more appropriate distribution for the LRT statistics when using a model selection criteria to determine the number of knots.

Last, in Chapter 4, we propose a penalized Cox proportional hazards (PH) model with functional covariates and interaction term, which can be used to examine the effect of time-varying exposures and their interaction effect on survival outcome. Flexible weighted cumulative exposure (WCE) model proposed by Sylvestre and Abrahamowicz (2009) and penalized scalar-on-functions regression model by Fuchs et al. (2015) gave insight into our model. Penalized Cox PH model discussed by Perperoglou (2014) and Gray (1992) provided methods of parameter estimation. Furthermore, parameter estimation can be implemented using standard statistical packages for survival analysis with time-dependent covariates making the proposed models applicable to many data applications (Therneau et al., 2003). Adequate performance of the proposed methods in both parameter estimation and hypothesis testing was demonstrated through simulation studies. The application of the proposed method in a cohort of depressed patients showed that there is non interaction between two classes of antidepressants on CAD events. The proposed model framework can be applied to other types of medication groups to examine the existence of interaction effects among medication groups as well as the effects of medication exposures.

In conclusion, with the availability of Electronic medical records (EMR), enormous quantities of clinical data, including medical diagnosis, laboratory testing, medication dispensing information, can be easily accessed by researchers and have been increasingly used in many

health systems around the country, which offers an unprecedented research opportunity for monitoring disease development, progression and treatment. In particular, medication dispensing data allow researchers to explore complex relationships among long-term medication use, disease progression and potential side-effects in large patient populations. Our flexible models can deal with medication exposures in several situations. More importantly, Our models not only provide novel solutions to solve problems in medical area, but also can be applied to many other fields when time-varying exposures are presented.



## BIBLIOGRAPHY

- Abrahamowicz, M., G. Bartlett, R. Tamblyn, and R. du Berger (2006). Modeling cumulative dose and exposure duration provided insights regarding the associations between benzodiazepines and injuries. *Journal of clinical epidemiology* 59(4), 393–403.
- Abrahamowicz, M., M.-E. Beauchamp, and M.-P. Sylvestre (2012). Comparison of alternative models for linking drug exposure with adverse effects. *Statistics in Medicine* 31(11-12), 1014–1030.
- Abrahamowicz, M., T. Mackenzie, and J. M. Esdaile (1996). Time-dependent hazard ratio: modeling and hypothesis testing with application in lupus nephritis. *Journal of the American Statistical Association* 91(436), 1432–1439.
- Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on* 19(6), 716–723.
- Andersen, P. K. and R. D. Gill (1982). Cox’s regression model for counting processes: a large sample study. *The annals of statistics*, 1100–1120.
- Bergstrom, R., L. Lemberger, N. Farid, and R. Wolen (1988). Clinical pharmacology and pharmacokinetics of fluoxetine: A review. *The British Journal of Psychiatry*.
- Berhane, K., M. Hauptmann, and B. Langholz (2008). Using tensor product splines in modeling exposure–time–response relationships: Application to the colorado plateau uranium miners cohort. *Statistics in medicine* 27(26), 5484–5496.
- Bhadra, D., M. J. Daniels, S. Kim, M. Ghosh, and B. Mukherjee (2012). A bayesian semiparametric approach for incorporating longitudinal information on exposure history for inference in case–control studies. *Biometrics* 68(2), 361–370.

- Breslow, N. E., J. Lubin, P. Marek, and B. Langholz (1983). Multiplicative models and cohort analysis. *Journal of the American Statistical Association* 78(381), 1–12.
- Buja, A., T. Hastie, and R. Tibshirani (1989). Linear smoothers and additive models. *The Annals of Statistics*, 453–510.
- Callahan, C. M., S. L. Hui, N. A. Nienaber, and B. S. Musick (1994). Longitudinal study of depression and health services use among elderly primary care patients. *Journal of the American Geriatrics Society*.
- Damush, T. M., H. Jia, L. D. Ried, H. Qin, R. Cameon, L. Plue, and L. S. Williams (2008). Case-finding algorithm for post-stroke depression in the veterans health administration. *International journal of geriatric psychiatry* 23(5), 517–522.
- De Boor, C. (2001). A practical guide to splines (rev. ed). *Applied Mathematical Sciences* 27.
- de la Torre, B. R., J. Dreher, I. Malevany, M. Bagli, M. Kolbinger, H. Omran, B. Lüderitz, and M. L. Rao (2001). Serum levels and cardiovascular effects of tricyclic antidepressants and selective serotonin reuptake inhibitors in depressed patients. *Therapeutic drug monitoring* 23(4), 435–440.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The annals of Statistics*, 1–26.
- Ferraty, F. and P. Vieu (2009). Additive prediction and boosting for functional data. *Computational Statistics & Data Analysis* 53(4), 1400–1413.
- Fuchs, K., F. Scheipl, and S. Greven (2015). Penalized scalar-on-functions regression with interaction term. *Computational Statistics & Data Analysis* 81, 38–51.
- Gasparrini, A. (2013). Modeling exposure–lag–response associations with distributed lag non-linear models. *Statistics in medicine*.

- Giorgi, R., M. Abrahamowicz, C. Quantin, P. Bolard, J. Esteve, J. Gouvernet, and J. Faivre (2003). A relative survival regression model using b-spline functions to model non-proportional hazards. *Statistics in medicine* 22(17), 2767–2784.
- Glassman, A. H. (1984). Cardiovascular effects of tricyclic antidepressants. *Annual review of medicine* 35(1), 503–511.
- Goldsmith, J., C. M. Crainiceanu, B. Caffo, and D. Reich (2012). Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 61(3), 453–469.
- Gray, R. J. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association* 87(420), 942–951.
- Gray, R. J. (1994). Spline-based tests in survival analysis. *Biometrics*, 640–652.
- Guo, X. and B. P. Carlin (2004). Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician* 58(1), 16–24.
- Hastie, T. and R. Tibshirani (1986). Generalized additive models. *Statistical science*, 297–310.
- Hastie, T. J. and R. J. Tibshirani (1990). *Generalized additive models*, Volume 43. CRC Press.
- Hauptmann, M., J. Wellmann, J. H. Lubin, P. S. Rosenberg, and L. Kreienbrock (2000). Analysis of exposure-time-response relationships using a spline weight function. *Biometrics* 56(4), 1105–1108.

- Hurvich, C. M., J. S. Simonoff, and C.-L. Tsai (1998). Smoothing parameter selection in nonparametric regression using an improved akaike information criterion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 60(2), 271–293.
- Hurvich, C. M. and C. Tsai (1990). The impact of model selection on inference in linear regression. *The American Statistician* 44(3), 214–217.
- Jefferson, J. W. (1975). A review of the cardiovascular effects and toxicity of tricyclic antidepressants. *Psychosomatic Medicine* 37(2), 160–179.
- Kohn, R., C. F. Ansley, and D. Tharm (1991). The performance of cross-validation and maximum likelihood estimators of spline smoothing parameters. *Journal of the American Statistical Association* 86(416), 1042–1050.
- Langholz, B., D. Thomas, A. Xiang, and D. Stram (1999). Latency analysis in epidemiologic studies of occupational exposures: application to the colorado plateau uranium miners cohort. *American journal of industrial medicine* 35(3), 246–256.
- Lindsley, C. W. (2012). The top prescription drugs of 2011 in the united states: antipsychotics and antidepressants once again lead cns therapeutics. *ACS chemical neuroscience* 3(8), 630–631.
- Linhart, H. and W. Zucchini (1986). *Model selection*. Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley.
- Liu, H., W. Tu, et al. (2012). A semiparametric regression model for paired longitudinal outcomes with application in childhood blood pressure development. *The Annals of Applied Statistics* 6(4), 1861–1882.
- Mahmud, M., M. Abrahamowicz, K. Leffondré, and Y. P. Chaubey (2006). Selecting the optimal transformation of a continuous covariate in cox’s regression: Implications for

- hypothesis testing. *Communications in Statistics Simulation and Computation*® 35(1), 27–45.
- McDonald, C. J., J. M. Overhage, W. M. Tierney, P. R. Dexter, D. K. Martin, J. G. Suico, A. Zafar, G. Schadow, L. Blevins, T. Glazener, et al. (1999). The regenstrief medical record system: a quarter century experience. *International journal of medical informatics* 54(3), 225–253.
- Pacher, P. and V. Kecskemeti (2004). Cardiovascular side effects of new antidepressants and antipsychotics: new drugs, old concerns? *Current pharmaceutical design* 10(20), 2463.
- Pacher, P., Z. Ungvari, P. P. Nanasi, S. Furst, and V. Kecskemeti (1999). Speculations on difference between tricyclic and selective serotonin reuptake inhibitor antidepressants on their cardiac effects. is there any? *Current medicinal chemistry* 6(6), 469–480.
- Perperoglou, A. (2014). Cox models with dynamic ridge penalties on time-varying effects of the covariates. *Statistics in medicine* 33(1), 170–180.
- Richardson, D. B. (2009). Latency models for analyses of protracted exposures. *Epidemiology (Cambridge, Mass.)* 20(3), 395.
- Robinson, G. K. (1991). That blup is a good thing: The estimation of random effects. *Statistical Science*, 15–32.
- Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of computational and graphical statistics* 11(4).
- Ruppert, D., M. P. Wand, and R. J. Carroll (2003). *Semiparametric regression*. Number 12. Cambridge University Press.

- Schipper, M., J. M. Taylor, and X. Lin (2007). Bayesian generalized monotonic functional mixed models for the effects of radiation dose histograms on normal tissue complications. *Statistics in medicine* 26(25), 4643–4656.
- Schipper, M., J. M. Taylor, and X. Lin (2008). Generalized monotonic functional mixed models with application to modelling normal tissue complications. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 57(2), 149–163.
- Searle, S. R., G. Casella, and C. E. McCulloch (2009). *Variance components*, Volume 391. John Wiley & Sons.
- Smitten, A. L., H. K. Choi, M. C. Hochberg, S. Suissa, T. A. Simon, M. A. Testa, and K. A. Chan (2008). The risk of hospitalized infection in patients with rheumatoid arthritis. *The Journal of rheumatology* 35(3), 387–393.
- Smoller, J. W., M. Allison, B. B. Cochrane, J. D. Curb, R. H. Perlis, J. G. Robinson, M. C. Rosal, N. K. Wenger, and S. Wassertheil-Smoller (2009). Antidepressant use and risk of incident cardiovascular morbidity and mortality among postmenopausal women in the women’s health initiative study. *Archives of internal medicine* 169(22), 2128.
- Stranges, S., M. R. Bonner, F. Fucci, K. M. Cummings, J. L. Freudenheim, J. M. Dorn, P. Muti, G. A. Giovino, A. Hyland, and M. Trevisan (2006). Lifetime cumulative exposure to secondhand smoke and risk of myocardial infarction in never smokers: results from the western new york health study, 1995-2001. *Archives of internal medicine* 166(18), 1961.
- Suissa, S. (2007). Immortal time bias in observational studies of drug effects. *Pharmacoepidemiology and drug safety* 16(3), 241–249.
- Sylvestre, M.-P. and M. Abrahamowicz (2008). Comparison of algorithms to generate event times conditional on time-dependent covariates. *Statistics in Medicine* 27(14), 2618–2634.

- Sylvestre, M.-P. and M. Abrahamowicz (2009). Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. *Statistics in medicine* 28(27), 3437–3453.
- Therneau, T. M., P. M. Grambsch, and V. S. Pankratz (2003). Penalized survival models and frailty. *Journal of computational and graphical statistics* 12(1), 156–175.
- Thomas, D. C. (1988). Models for exposure-time-response relationships with applications to cancer epidemiology. *Annual review of public health* 9(1), 451–482.
- VACEK, P. M. (1997). Assessing the effect of intensity when exposure varies over time. *Statistics in medicine* 16(5), 505–513.
- Wahba, G. (1983). Bayesian” confidence intervals” for the cross-validated smoothing spline. *Journal of the Royal Statistical Society. Series B (Methodological)*, 133–150.
- Wolfe, F., L. Caplan, and K. Michaud (2006). Treatment for rheumatoid arthritis and the risk of hospitalization for pneumonia: Associations with prednisone, disease-modifying antirheumatic drugs, and anti-tumor necrosis factor therapy. *Arthritis & Rheumatism* 54(2), 628–634.
- Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73(1), 3–36.
- Zellweger, M. J., R. H. Osterwalder, W. Langewitz, and M. E. Pfisterer (2004). Coronary artery disease and depression. *European Heart Journal* 25(1), 3–9.
- Zhang, D., X. Lin, and M. Sowers (2007). Two-stage functional mixed models for evaluating the effect of longitudinal covariate profiles on a scalar outcome. *Biometrics* 63(2), 351–362.

# CURRICULUM VITAE

Chenkun Wang

## EDUCATION

- Ph.D. in Biostatistics, Indiana University, Indianapolis, IN, 2015 (Minor in Epidemiology)
- M.S. in Probability and Mathematical Statistics, Nanjing University, Nanjing, China, 2009
- B.S. in Computational Mathematics, Anhui University, Hefei, China, 2006

## WORKING EXPERIENCE

- Biostatistics Intern, Vertex Pharmaceuticals Inc, Boston, MA, 05/2014 - 08/2014
- Research Assistant, Department of Biostatistics, Indiana University School of Medicine, Richard M. Fairbank School of Public Health, Indianapolis, IN, 07/2011 - 03/2015